Proba Stats - Automne 2020

Contents

0 Bref rappel de la théorie de la mesure

Définition 1

Soit E un ensemble. Une tribu (ou σ-algèbre) sur E est une famille $\mathcal{A}$ de parties de E satisfaisant:

i) $E \in \mathcal{A}$

ii) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$

iii) Si $A_n \in \mathcal{A}$ $\forall n \in \mathbb{N}$, alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

Terminologie: $A \in \mathcal{A}$ est une partie ($\mathcal{A}$-)mesurable, $(E, \mathcal{A})$ est un espace mesurable.

Définition 2

Soit $\mathcal{C} \subset \mathcal{P}(E)$. Alors $\sigma(\mathcal{C}) := \bigcap_{\substack{\mathcal{A} \text{ tribu} \\ \mathcal{C} \subset \mathcal{A}}} \mathcal{A}$

Exemple 3

i) E un espace topologique avec une famille d'ouverts $\mathcal{O}$.

$$\mathcal{B}(E) := \sigma(\mathcal{O}) = \quad \text{tribu borélienne de } E$$

ii) Soit $(E_1, \mathcal{A}_1)$ et $(E_2, \mathcal{A}_2)$ des espaces mesurables. La tribu produit sur $E_1 \times E_2$ est:

$$\mathcal{A}_1 \otimes A_2 := \sigma(A_1 \times A_2 \mid A_i \in \mathcal{A}_i)$$

Définition 4

Une mesure (positive) sur $(E, \mathcal{A})$ est une fonction $\mu : \mathcal{A} \to [0, \infty]$ satisfaisant $\mu(\varnothing) = 0$ et $\mu\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$ où $(A_n)_{n \in \mathbb{N}}$ est une famille de parties mesurables disjointes.

Exemple 5

i) E fini ou dénombrable, $\mathcal{A} = \mathcal{P}(E)$ la mesure de comptage.

$$\mu(A) := |A|$$

ii) Pour $x \in E$ on définit la mesure de dirac en x par:

$$\delta_x(A) := \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

iii) La mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, l'unique mesure λ satisfaisant $\lambda\big((a, b)\big) = b - a \quad \forall a < b$.

Définition 6

Soit $(E, \mathcal{A})$, $(F, \mathcal{B})$ deux espaces mesurables. Une fonction $f : E \to F$ est appellée mesurable si $\forall B \in \mathcal{B}$, $f^{-1}(B) \in \mathcal{A}$

Définition 7

Soit $(E,\mathcal{A})$, $(F,\mathcal{B})$ deux espaces mesurables, μ une mesure sur $(E,\mathcal{A})$, $f: E \to F$ mesurable. $f_*\mu$ (parfois aussi $f(\mu)$) est la mesure (mesure-image) sur $(F,\mathcal{B})$ définie $\forall B \in \mathcal{B}$:

$$f_*\mu(B) := \mu\left(f^{-1}(B)\right)$$

Soit μ une mesure sur $(E,\mathcal{A})$ et $f: E \to [0,\infty]$. On note par $\int f\, d\mu = \int f(x)\, \mu(dx)\ \left(= \int f(x)\, d\mu(x)\right) \in [0,\infty]$ l'intégrale de f par rapport à μ.

$f: E \to \mathbb{R}$ est intégrable si $\int |f|\, d\mu < \infty$, et dans ce cas

$$\int f\, d\mu := \int f_+\, d\mu - \int f_-\, d\mu \quad \text{où} \quad f_+ := \max(f,0) \quad \text{et} \quad f_- := \max(-f,0)$$

Mesure produit

Soient μ et ν deux mesures σ-finies sur $(E,\mathcal{A})$ et $(F,\mathcal{B})$ respectivement.
Il existe une unique mesure, notée $\mu \otimes \nu$, sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ telle que $\forall A \in \mathcal{A}$, $\forall B \in \mathcal{B}$:

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

Théorème de Fubini-Tonelli

Soit $f: E \times F \to [0,\infty]$ mesurable. Alors

$$\int\limits_{E\times F} f\, d(\mu \otimes \nu) = \int_E \left(\int_F f(x,y)\, \nu(dy)\right)\, \mu(dx) = \int_F \left(\int_E f(x,y)\, \mu(dy)\right)\, \nu(dx)$$

1 Fondaments de la théorie des probabilites

1.1 Espaces de probabilité

Motivation: Prenons un système aléatoire. Son état est un élément ω dans un ensemble Ω, l'ensemble des réalisations ou des déterminations du hasard.

Exemple A - Lancé de dé: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Exemple B - lancer de fléchettes: $\Omega = \overline{D} = \{\omega \in \mathbb{C} \mid |\omega| \leq 1\}$
$\mathcal{A} \subset \mathcal{P}(\Omega)$ est la famille des événements, c'est à dire les parties de Ω dont on peut évaluer les probabilités.

Exemple A $A = \{2, 4, 6\}$: j'ai lancé un nombre pair. $A = \{6\}$: j'ai lancé un 6.

Exemple B $A = 0.05 \cdot \overline{D} = \{\omega \in \Omega \mid |\omega| \leq 0.05\}$: j'ai tiré dans le bull's eye.

Une fonction $\mathbb{P} : \mathcal{A} \to [0, 1]$ où $\mathbb{P}(A)$ est la probabilité d'obtenir l'événement A.

Exemple A $\mathbb{P}(\{6\}) = \frac{1}{6}$, $\mathbb{P}(\{2, 4, 6\}) = \frac{1}{2}$ pour un dé équilibré.

Exemple B $\mathbb{P}(0.05 \cdot \overline{D}) = \frac{1}{9^2}$

Clairement on a $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(A \sqcup B) = \mathbb{P}(A) + \mathbb{P}(B)$ pour A et B disjoints.

Définition 1 (Kolmogarov, 1933)

Un espace de probabilité est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$, où $(\Omega, \mathcal{A})$ est un espace mesurable et $\mathbb{P}$ est une mesure sur $(\Omega, \mathcal{A})$ satisfaisant $\mathbb{P}(\Omega) = 1$. On appelle ces mesures les mesures de probabilités.

Exemple 2

On lance un dé deux fois: $\Omega = \{1, \ldots, 6\}^2$, $\mathcal{A} = \mathcal{P}(\Omega)$. Partons de l'hypothèse que le dé n'est pas biaisé, alors: $\mathbb{P}(A) = \frac{\#A}{36}$

Exemple 3

On lance un dé (équilibré) jusqu'à obtenir un 6. $\Omega = \{1, \ldots, 6\}^{\mathbb{N}^*}$. Notons que l'ensemble des réalisations est infini et que prendre comme tribu $\mathcal{P}(\Omega)$ nous donnerait une tribu bien trop grande. C'est pour ça que nous poserons:

$$\mathcal{A} := \sigma\left(\{\omega \in \Omega \mid \omega_1 = i_1,\ \omega_2 = i_2, \ldots,\ \omega_n = i_n,\ n \in \mathbb{N}^*, i_1, \ldots, i_n \in \{1, \ldots, 6\}\}\right)$$

Pour $\mathbb{P}$, on aimerait avoir comme propriété:

$$\mathbb{P}(\{\omega \in \Omega \mid \omega_1 = i_1, \ldots, \omega_n = i_n\}) = \left(\frac{1}{6}\right)^n \quad \text{pour tout } i_1, \ldots, i_n \in \{1, \ldots, 6\}$$

On verra qu'il existe un unique $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ satisfaisant cette propriété.

Exemple 4: Mouvement Brownien

Prenons une particule partant de $0 \in \mathbb{R}^3$ au temps $t_0 = 0$ avec une trajectoire continue et aléatoire. On peut alors poser:

$$\Omega := C(\mathbb{R}_+, \mathbb{R}^3) \quad \text{et} \quad \mathcal{A} := \sigma\big(\{\omega \in \Omega \mid \omega(t) \in B\} \mid t \in \mathbb{R}_+,\ B \in \mathcal{B}(\mathbb{R}^3)\big)$$

1.2 Variable aléatoire

Motivation: Nous voulons une variable X dont la valeur dépend de la réalisation ω.

Définition 5

Soit $(E,\mathcal{E})$ un espace mesurable. Une variable aléatoire (Noté: VA) (Souvent omis: à valeur) dans E est une fonction mesurable $X:\ \Omega \to E$

Remarque: Si $(E,\mathcal{E})$ n'est pas précisé alors il est sous-entendu: $(E,\mathcal{E}) := (\mathbb{R},\mathcal{B}(\mathbb{R}))$

Exemple 2 (suite)

Notons nos lancés: $\omega = (i,j) \in \{1,\dots,6\}^2$, on peut par exemple prendre la variable aléatoire: $X\big((i,j)\big) := i+j$, la somme des points.

Exemple 3 (suite)

$\omega = (\omega_i)_{i\in\mathbb{N}^*} \in \Omega$, on peut noter la variable aléatoire:

$$X(\omega) := \inf\{i \in \mathbb{N}^* \mid \omega_i = 6\} \quad \text{avec la convention} \quad \inf\varnothing = \infty$$

Notons alors que $X:\ \Omega \to \mathbb{N}^*\cup\{\infty\}$, vérifions que X soit bien mesurable, $\forall k \in \mathbb{N}^*$:

$$X^{-1}(\{k\}) = \{\omega \in \Omega \mid \omega_1 \neq 6, \dots, \omega_{k-1} \neq 6, \omega_k = 6\}$$

Exemple 4 (suite)

On pourait noter la variable aléatoire naturelle: $X(\omega) = \omega(t) \in \mathbb{R}^3$.

Définition 6

La loi (ou la distribution) de la VA X est la mesure $\mathbb{P}_X := X_*\mathbb{P}$ sur $(E,\mathcal{E})$ et donc $(E,\mathcal{E},\mathbb{P}_X)$ est un espace de proba. Et alors:

$$\forall B \in \mathcal{E},\ \mathbb{P}_X(B) = \mathbb{P}\left(X^{-1}(B)\right) = \mathbb{P}\big(\{\omega \mid X(\omega) \in B\}\big) =: \mathbb{P}(X \in B)$$

La notation probabiliste: $\{\omega \mid X(\omega \in B)\} =: \{X \in B\}$

Remarque "banale"

Soit μ une mesure de probabilité sur un espace mesurable $(E,\mathcal{E})$. Comment construire une VA avec loi μ ?

Posons juste l'identité:

$$(\Omega,\mathcal{A},\mathbb{P}) := (E,\mathcal{E},\mu) \quad \text{et} \quad X(\omega) := \omega$$

Cas particuliers "élémentaires"

- VA discrète.
  E fini ou dénombrable, $\mathcal{E} = \mathcal{P}(E)$. Alors $\mathbb{P}_X = \sum_{x\in E} p_x \cdot \delta_x$ où $p_x := \mathbb{P}(X=x)$.

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\left(\bigcup_{x \in B} \{X = x\}\right) \stackrel{*}{=} \sum_{x \in B} \mathbb{P}(X = x) = \sum_{x \in E} p_x \cdot \delta_x(B)$$

$(*)$ Notons que cette égalité est justifiée par le fait que B soit au plus dénombrable et que l'union est disjointe.

$\lrcorner$

Exemple 3 (suite)

$X \in \mathbb{N}^* \cup \{\infty\}$. Pour $k \geq 1$, on a:

$$\mathbb{P}(X = k) = \mathbb{P}(\omega_1 \neq 6,\ \omega_2 \neq 6, \ldots, \omega_{k-1} \neq 6,\ \omega_k = 6) = \mathbb{P}\left(\bigcup_{i_1, \ldots, i_{k-1} = 1}^{5} \{\omega_1 = i_1, \omega_2 = i_2, \ldots, \omega_{k-1} = i_{k-1}, \omega_k = 6\}\right)$$
$$= 5^{k-1} \left(\frac{1}{6}\right)^k = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}$$

Remarque: $\mathbb{P}(X = \infty) = 1 - \sum_{k=1}^{\infty} \mathbb{P}(X = k) = 0$

Notons néanmoins que $\{X = \infty\}$ contient énormément d'éléments de Ω. Il est non-dénombrable.

- VA à densité.
  $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue.
  Par le théorème de Radon-Nikodym, il existe une fonction mesurable $p : \mathbb{R}^d \to \mathbb{R}_+$ telle que:

$$\mathbb{P}_X(B) = \int_B p(x)\, dx$$

Terminologie: p est appelé la densité (de la loi) de X.

1.3 Espérance mathématique

Définition 7

Soit X une VA réelle. L'espérance de X est:

$$\mathbb{E}[X] := \int X(\omega)\, \mathbb{P}(d\omega)$$

qui est bien définie si $X \geq 0$, (alors $\mathbb{E}[X] \in [0, \infty]$) ou $\mathbb{E}[|X|] < \infty$ (X est intégrable).

Si $X = (X_1, \ldots, X_n) \in \mathbb{R}^d$, on définit:

$$\mathbb{E}[X] := (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$$

Proposition 8

Pour une VA X dans $(E, \mathcal{E})$ et $f:\, E \to \mathbb{R} \cup \{\infty\}$ mesurable, on a:

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

si $f \geq 0$ ou si $\mathbb{E}[|f(X)|] < \infty$

Démonstration: Exercice 1.1

Exemple 2 (suite)

Lancé de deux dés et la variable aléatoire est la somme des résultats.

$$\mathbb{E}[X] = \frac{1}{36} \sum_{i,j=1}^{6} (i + j) = \cdots = 7$$

Exemple 3 (suite)

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) = \frac{1}{6} \sum_{k=1}^{\infty} k \cdot \left(\frac{5}{6}\right)^{k-1}$$

Notons l'identité $\displaystyle\sum_{k=1}^{\infty} k \cdot a^{k-1} = \frac{1}{(1-a)^2}$ et alors $\mathbb{E}[X] = 6$

1.4 Lois classiques

Lois discrètes

(a) Loi uniforme.
E fini, et $\forall x \in E$
$$\mathbb{P}(X = x) = \frac{1}{\#E}$$

(b) Loi de Bernoulli. ($p \in [0, 1]$)
$E = \{0, 1\}$
$$\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p$$
Interprétation: pièce truquée

(c) Loi binômiale. ($n \in \mathbb{N}^*,\ p \in [0, 1]$)
$E = \{0, \ldots, n\}$
$$\mathbb{P}(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$
Interprétation: nombre de faces obtenus dans n lancés d'une pièce truquée

(d) Loi géométrique. ($p \in (0, 1)$)
$E = \mathbb{N}$
$$\mathbb{P}(X = k) = (1 - p) \cdot p^k$$
Interprétation: nombre de faces avant le premier "pile"

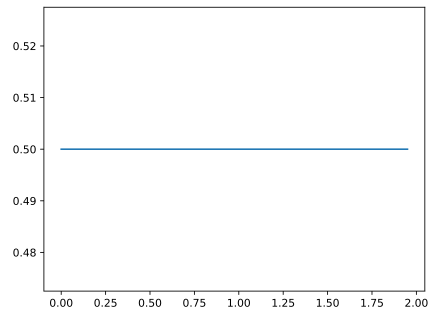
(e) Loi de Poisson. ($\lambda \in \mathbb{R}_+$)
$E = \mathbb{N}$
$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$
Interprétation: Nombre d'événements rares durant une longue période X_n a loi binômiale avec paramètres n et p_n. Si $n \cdot p_n \underset{n\to\infty}{\longrightarrow} \lambda$ alors $\mathbb{P}(X_n = k) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X = k)$ (de Poisson)

Lois continues

X est VA dans $\mathbb{R}$ avec densité p.

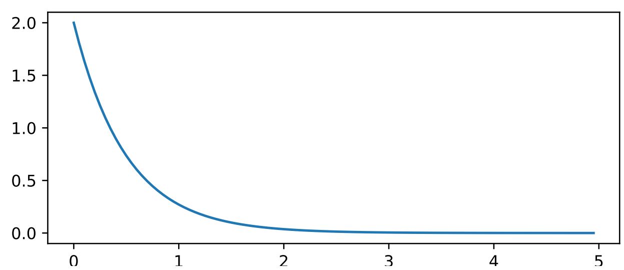
(a) Loi uniforme sur $[a, b]$

$$p(x) = \frac{1}{b-a} \cdot 1_{[a,b]}(x)$$



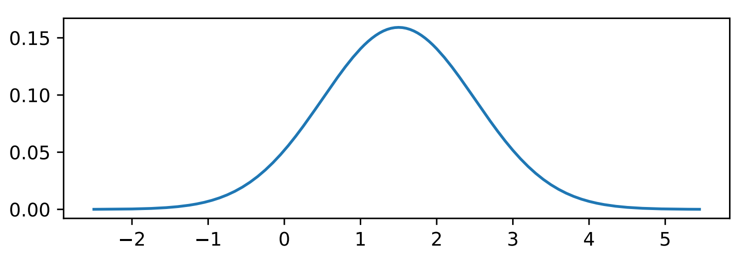
(b) Loi exponentielle ($\lambda > 0$)

$$p(x) = 1_{\mathbb{R}_+}(x) \cdot \lambda \cdot e^{-\lambda x}$$



(c) Loi Gaussienne (ou normale) ($m \in \mathbb{R}, \sigma > 0$)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$



1.5 Fonction de répartition

Définition 9

Pour une VA réelle X, on définit la fonction de répartition: $F_X : \mathbb{R} \to [0,1]$ par:

$$F_X(t) := \mathbb{P}(X \leq t)$$

Proposition 10

Si $F = F_X$ est la fonction de répartition d'une VA X, alors:

(i) F est croissante.

(ii) $\lim\limits_{t\to-\infty} F(t) = 0$, et $\lim\limits_{t\to\infty} F(t) = 1$

(iii) F est continue par à droite: $\lim\limits_{s\searrow t} F(s) = F(t)$

<u>*Démonstration:*</u>

(i) évident.

(ii) Avec $A_n = \{X \leq n\}$ et $B_n = \{X \leq -n\}$ on a $\Omega = \bigcup\limits_{n\in\mathbb{N}} A_n$ et $\varnothing = \bigcap\limits_{n\in\mathbb{N}} B_n$

En utilisant l'exercice 1.2, par exemple:

$$1 = \mathbb{P}(\Omega) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} F(n)$$

(iii) Soit $(a_n)_{n\in\mathbb{N}}$ une suite décroissante satisfaisant $a_n > 0$ $\forall n$ et $a_n \to 0$. Avec $A_n = \{X \leq t + a_n\}$ on a

$$\lim_{n\to\infty} F(t+a_n) = \lim_{n\to\infty} \mathbb{P}(A_n) \overset{\text{ex } 1.2}{=} \mathbb{P}\left(\bigcap_n A_n\right) = \mathbb{P}(X \leq t)$$

$\square$

Inversément, est-ce que chaque fonction F satisfaisant ces propriétés est une fonction de répartition ? (voui)

Proposition 11

Si F satisfait (i)-(iii) de la proposition 10 alors il existe une VA telle que $F = F_X$

<u>*Démonstration:*</u>
Soit $\Omega := (0,1)$, $\mathcal{A} = \mathcal{B}\big((0,1)\big)$, $\mathbb{P} =$ mesure Lebesgue. Pour $\omega \in \Omega$, on définit:

$$X(\omega) := \sup\{s \mid F(s) < \omega\}$$

Nous allons montrer que, $\forall t \in \mathbb{R}$: $\{\omega \mid X(\omega) \leq t\} = \{\omega \mid \omega \leq F(t)\}$
Supposons que nous l'avons montré, alors:

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}\big(\{\omega \mid \omega \leq F(f)\}\big) = F(t)$$

Montrons maintenant l'identité, par double inclusion. On observe:

- $\omega \leq F(t) \implies t \geq X(\omega)$ (Par definition de X et du sup)

- $\omega > F(t) \overset{F \text{ cont. droite}}{\implies} \exists \varepsilon > 0,\ F(t+\varepsilon) < \omega \implies X(\omega) \geq t + \varepsilon \implies X(\omega) > t$

$\square$

Remarque 12: La fonction $X : (0,1) \to \mathbb{R}$ construite en haut est souvent appelée inverse de F, notée F^{-1}, même si F n'est pas bijective.

1.6 Tribu engendrée par une variable aléatoire

Définition 13

Soit X une VA à valeurs dans $(E, \mathcal{E})$. La tribu engendrée par X est:

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

(La plus petite tribu sur Ω qui rende X mesurable)

Proposition 14

Soit X une VA à valeurs dans $(E, \mathcal{E})$ et Y une VA à valeurs réelles. Sont équivalents:

(i) Y est $\sigma(X)$-mesurable

(ii) $\exists f : (E, \mathcal{E}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ mesurable telle que $Y = f(X)$

Démonstration:

(ii) $\Longrightarrow$ (i) Exercice

(i) $\Longrightarrow$ (ii) Soit Y $\sigma(X)$-mesurable.

Cas 1: Y étagée. $Y = \sum_{i=1}^{n} \lambda_i 1_{A_i}$, $A_i \in \sigma(X)$ Alors $\forall i$, $\exists B_i \in \mathcal{E}$ tel que $A_i = X^{-1}(B_i)$, et donc:

$$Y = \sum_{i=1}^{n} \lambda_i 1_{X^{-1}(B_i)} = \sum_{i=1}^{n} \lambda_i 1_{b_i} \circ X = f \circ X$$

où $f := \sum_{i=1}^{n} \lambda_i 1_{B_i}$ mesurable.

Cas 2: Y général.
On a $Y = \lim_{n\to\infty} Y_n$, où Y_n sont étagées et Y_n sont $\sigma(X)$-mesurables. Par le cas 1, $\forall n$, $\exists f_n : E \to \mathbb{R}$ mesurable telle que $Y_n = f_n(X)$. On définit:

$$f(x) := \begin{cases} \lim_{n\to\infty} f_n(x) & \text{si la limite existe} \\ 0 & \text{sinon} \end{cases}$$

f est mesurable.
De plus, $\forall \omega \in \Omega$, la limite $\lim_{n\to\infty} f_n(x)$ existe pour $x = X(\omega)$. (puisque $f_n(X(\omega)) = Y_n(\omega) \to Y(\omega)$)
Finalement, $f\big(X(\omega)\big) = \lim_{n\to\infty} f_n\big(X(\omega)\big) = Y(\omega)$

$\square$

1.7 Moments et inégalités de base

Définition 15

Soit X une VA réelle et $\mathbb{R} \ni p \geq 1$. Le moment d'ordre p de X est $\mathbb{E}[X^p]$ définie si $X \geq 0$ ou $p \in \mathbb{N}^*$ et $\mathbb{E}[|X|^p] < \infty$ sinon.

Nous parlerons d'espaces $L^p \equiv L^p(\Omega, \mathcal{A}, \mathbb{P})$, $p \in [1, \infty]$ sont définie comme dans le cours d'analyse fonctionnelle, avec la norme:

$$\|X\|_p := (\mathbb{E}[|X|^p])^{\frac{1}{p}} \quad \text{où } p < \infty, \quad \|X\|_\infty := \text{ess sup}\,|X| = \inf\{t \geq 0 \mid \mathbb{P}(X > t) = 0\}$$

En probabilité, on écrit presque sûrement (p.s.) à la place du "presque partout" que l'on rencontre en analyse.

Proposition 16: Inégalité de Hölder

$1 \leq p,\ q \leq \infty,\ \dfrac{1}{p} + \dfrac{1}{q} = 1$, alors:

$$\|XY\|_1 \leq \|X\|_p \cdot \|Y\|_q$$

Cas particuliers importants

(i) $\|X\|_p \leq \|X\|_q$ si $1 \leq p \leq q \leq \infty$

(ii) $\mathbb{E}[\,|XY|\,] \leq \mathbb{E}[\,X^2\,]^{\frac{1}{2}} \cdot \mathbb{E}[\,Y^2\,]^{\frac{1}{2}}$ (Cauchy-Schwarz)

(iii) $\mathbb{E}[\,|X|\,]^2 \leq \mathbb{E}[X^2]$

Définition 17

Soit $X \in L^2$. La variance de X est

$$\text{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Remarque 18

- $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- $\forall a \in \mathbb{R}$, $\mathbb{E}[(X-a)]^2 = \text{Var}(X) + (\mathbb{E}[X] - a)^2$ et donc $\text{Var}(X) = \inf\limits_{a \in \mathbb{R}} \mathbb{E}\left[(X-a)^2\right]$

Proposition 19:(Bienaymé-Tchebychev-Markov)

Soit $f : \mathbb{R} \to \mathbb{R}_+$ croissante et X une VA réelle. Alors pour tout $a \in \mathbb{R}$ on a

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

$\ulcorner$

$$\mathbb{P}(X \geq a) \leq \mathbb{P}\big(f(x) \geq f(a)\big) = \mathbb{E}\left[1_{f(x) \geq f(a)}\right] \leq \mathbb{E}\left[1_{f(x) \geq f(a)} \cdot \frac{f(x)}{f(a)}\right] \leq \frac{\mathbb{E}[f(x)]}{f(a)}$$

$\lrcorner$

Cas Particuliers importants:

- $X \geq 0$, $a > 0$, $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}[X]}{a}$ (Markov)
- $X \in L^2$, $a \geq 0$, $\mathbb{P}\left(|X - \mathbb{E}[X]| > a\right) \leq \dfrac{\mathrm{Var}(X)}{a^2}$ (Tchebychev)
- $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}[e^{\lambda \cdot X}}{e^{a \cdot \lambda}}$ et ce $\forall \lambda > 0$ (Chernov)

Définition 20

(i) $X, Y \in L^2$, on défini la covariance:

$$\mathrm{Cov}(X,Y) := \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

(ii) $X = (X_1, \ldots, X_d)$ avec $X_i \in L^2$ $\forall i$

$$\mathrm{Cov}(X) = \big(\mathrm{Cov}(X_i, X_j)\big)_{i,j=1}^{d}$$

C'est donc une matrice $d \times d$

Remarque 21 $\mathrm{Cov}(X)$ est une matrice positive, i.e.

$$< \lambda,\ \mathrm{Cov}(X)\lambda >\geq 0 \quad \forall \lambda \in \mathbb{R}^d$$

2 Indépendance

2.1 Évènements indépendants

Définition 1

Deux évènements dans la tribu $A, B \in \mathcal{A}$ sont indépendants si

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

Interpretation: C'est le fait de savoir que B est réalisé ne donne pas d'information sur si A s'est réalisé ou pas. Plus formellement on peut parler de probabilité conditionnelle. $\mathbb{P}(A \mid B)$ est lu: Probabilité de A sachant que B est réalisé.

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

Mais alors si A et B sont indépendants nous avons: $\mathbb{P}(A \mid B) = \mathbb{P}(A)$

Exemple 2

(i) Lancer de deux dés.
$A = \{6\} \times \{1, \ldots, 6\}$ et $B = \{1, \ldots, 6\} \times \{6\}$ sont indépendants.

(ii) Lancer d'un seul dé. $A = \{1, 2\}$, $B = \{1, 3, 5\}$ sont indépendants.
En effet on vérifie facilement:

$$\mathbb{P}(A \cap B) = \frac{1}{6} = \frac{2}{6} \cdot \frac{3}{6} = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Définition 3

Une famille d'évènements $\{A_i\}_{i \in I}$ est indépendants si pour tout $J \subset I$ fini on a:

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

Remarque 4 Soit I fini. Il ne suffit pas que

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$$

Aussi, il ne suffit pas que pour chaque pair $\{i, j\} \subset I$, A_i et A_j sont indépendants. Il faut bien que l'égalité tienne pour tout les sous ensemble J de I.

Deux lancers de pièce équilibrées.
$A = \{1\} \times \{0, 1\}$, $\mathbb{P}(A) = \frac{1}{2}$, $B = \{0, 1\} \times \{1\}$, $\mathbb{P}(B) = \frac{1}{2}$ et $C = \big\{(0,0),\ (1,1)\big\}$, $\mathbb{P}(C) = \frac{1}{4}$
$\mathbb{P}(A \cap B) = \frac{1}{4}$, $\mathbb{P}(A \cap C) = \frac{1}{4}$, $\mathbb{P}(B \cap C) = \frac{1}{4}$, $\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8}$

2.2 Lemme de classe monotone (intermezzo)

C'est une excursion en théorie de la mesure.
Soit E un ensemble.

Définition 5

$\mathcal{M} \subset \mathcal{P}(E)$ est une classe monotone si

(i) $E \in \mathcal{M}$

(ii) $A, B \in \mathcal{M}$ et $A \subset B \implies B \setminus A \in \mathcal{M}$

(iii) $A_n \in \mathcal{M}$ où $A_n \subset A_{n+1} \implies \bigcup\limits_n A_n \in \mathcal{M}$

Définition 6

La classe monotone engendrée par une famille $\mathcal{C} \subset \mathcal{P}(E)$ est:

$$\mathcal{M}(\mathcal{C}) := \bigcap_{\substack{\mathcal{M} \supset \mathcal{C} \\ \mathcal{M} \text{ class. mono.}}} \mathcal{M}$$

Proposition 7: Lemme de classe monotone

Si $\mathcal{C} \subset \mathcal{P}(E)$ est stable par intersection finies ($A, B \in \mathcal{C} \implies A \cap B \in \mathcal{C}$) alors:

$$\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$$

<u>*Démonstration:*</u>
Une tribu est une classe monotone $\implies \mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$. Pour montrer que $\sigma(\mathcal{C}) \subset \mathcal{M}(\mathcal{C})$ on montre que $\mathcal{M}(\mathcal{C})$ est une tribu.
D'abord on trouve que une classe monotone $\mathcal{M}$ est une tribu si et seulement si elle est stable par intersection finie.

$\mathcal{M}$ classe monotone stable par intersection finie $\implies A, B \in \mathcal{M} \implies A^c, B^c \in \mathcal{M}$
$\implies A^c \cap B^c \in \mathcal{M} \implies A \cup B \in \mathcal{M}$ Et donc $\mathcal{M}$ est stable par unions finies.
Soient $A_1, A_2, \dots \in \mathcal{M}$ et posons $B_n := A_1 \cup \dots \cup A_n$. Alors $B_n \in \mathcal{M}$ par le resultat précédent. Et alors:

$$\bigcup_n A_n = \bigcup_n B_n \in \mathcal{M}$$

Il suffit de montrer que $A, B \in \mathcal{M}(\mathcal{C}) \implies A \cap B \in \mathcal{M}(\mathcal{C})$.
Soit $A \in \mathcal{C}$ fixé. Posons:

$$\mathcal{M}_1 := \{B \in \mathcal{M}(\mathcal{C}) \mid A \cap B \in \mathcal{M}(\mathcal{C})\}$$

$\mathcal{C}$ est stable par intersections finies $\implies$ (a) $\mathcal{C} \subset \mathcal{M}_1$
De plus, (b) $\mathcal{M}_1$ est une classe monotone.

(i): trivial
(ii): $B, B' \in \mathcal{M}_1$ et $B \subset B'$ nous avons alors:

$$\implies A \cap (B' \setminus B) = (A \cap B') \setminus (A \cap B) \in \mathcal{M}(\mathcal{C}) \implies B' \setminus B \in \mathcal{M}_1$$

(iii): $B_n \in \mathcal{M}_1$ et $B_n \subset B_{n+1}$

$$A \cap \left(\bigcup_n B_n \right) = \bigcup_n \underbrace{(A \cap B_n)}_{\in \mathcal{M}(\mathcal{C})} \in \mathcal{M}(\mathcal{C}) \implies \bigcup_n B_n \in \mathcal{M}$$

$\lrcorner$

Et alors (a) et (b) $\implies \mathcal{M}(\mathcal{C}) \subset \mathcal{M}_1 \implies \forall A \in \mathcal{C},\ \forall B \in \mathcal{M}(\mathcal{C}),\ A \cap B \in \mathcal{M}(\mathcal{C})$

Maintenant on répète cet argument en fixant $B \in \mathcal{M}(\mathcal{C})$ et on définit:

$$\mathcal{M}_2 := \{A \in \mathcal{M}(\mathcal{C}) \mid A \cap B \in \mathcal{M}(\mathcal{C})\}$$

On a montré que $\mathcal{C} \subset \mathcal{M}_2$. Comme en haut, $\mathcal{M}_2$ est une classe monotone, donc $\mathcal{M}(\mathcal{C}) \subset \mathcal{M}_2$

$\square$

Corollaire 8

Soient μ et ν deux mesures de probabilité sur $(E,\ \mathcal{A})$. S'il existe une famille $\mathcal{C} \subset \mathcal{A}$ stable par intersections finies telle que $\sigma(\mathcal{C}) = \mathcal{A}$ et $\mu(A) = \nu(A)\ \forall A \in \mathcal{C}$, alors $\mu = \nu$.

Démonstration:
Posons:

$$G := \{A \in \mathcal{A} \mid \mu(A) = \nu(A)\}$$

Alors $\mathcal{C} \subset G$ et G est une classe monotone (il suffit de verifier les axiomes, banalité). De plus par la proposition 7:

$$\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C}) = \mathcal{A} \quad \text{et on a} \quad \mathcal{A} = \mathcal{M}(\mathcal{C}) \subset G \subset \mathcal{A}$$

$\square$

Exemple 9: Unicité de la mesure de Lebesgue

Il existe au plus une mesure λ sur $\big([0,1],\ \mathcal{B}([0,1])\big)$ satisfaisant $\lambda(\ (a,b)\ ) = b - a\ \ \forall\ 0 \leq a < b \leq 1$
Prendre $\mathcal{C} := \{(a,b) \mid 0 \leq a < b \leq 1\}$

2.3 Variables aléatoires et tribus indépendantes

Définition 10

(i) Soient $\mathcal{B}_1, \dots, \mathcal{B}_n$ des sous-tribus de $\mathcal{A}$. $\mathcal{B}_1, \dots, \mathcal{B}_n$ sont indépendantes si:

$$\forall A_1 \in \mathcal{B}_1, \dots, A_n \in \mathcal{B}_n \quad \text{on a} \quad \mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$$

(ii) Soient $X_1, \dots, X_n$ des VA. $X_1, \dots, X_n$ sont indépendantes si $\sigma(X_1), \dots, \sigma(X_n)$ le sont.
(Si X_i est à valeurs dans $(E_i, \mathcal{E}_i)$, ceci veut dire que $\mathbb{P}(X_1 \in F_1, \dots, X_n \in F_n) = \mathbb{P}(X_i \in F_1) \cdots \mathbb{P}(X_n \in F_n)$ et ce $\forall F_1 \in \mathcal{E}_1, \dots, F_n \in \mathcal{E}_n$.)
On peut aussi noter alors:

Proposition 11

Les VA $X_1, \dots, X_n$ sont indépendantes si et seulement si la loi de $(X_1, \dots, X_n)$ est le produit des lois de $X_1, \dots, X_n$:

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

(égalité de mesures sur $(E_1 \times \dots \times E_n,\ \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n)$)
Dans ce cas on a:

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}\left[f_i(X_i)\right] \quad \text{pour tout } f_i \text{ positive et mesurable}$$

Démonstration:
Soient $F_1 \in \mathcal{E}_i,\ i = 1, \dots, n$. On a alors:

$$\mathbb{P}_{(X_1, \dots, X_n)}(F_1 \times \dots \times F_n) = \mathbb{P}\big((X_1, \dots, X_n) \in F_1 \times \dots \times F_n\big) = \mathbb{P}(X_1 \in F_1, \dots, X_n \in F_n)$$

et $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}(F_1 \times \dots \times F_n) = \mathbb{P}(X_1 \in F_1) \cdots \mathbb{P}(X_n \in F_n)$.
Par la remarque (ii) de la définition 10, on conclut que $X_1, \dots, X_n$ sont indépendantes si et seulement si $\mathbb{P}_{X_1, \dots, X_n}$ et $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ coincident sur tous les pavés $F_1 \times \dots \times F_n,\ F_i \in \mathcal{E}_i$.
La famille des pavés $\mathcal{C} = \{F_1 \times \dots \times F_n \mid F_i \in \mathcal{E}_i \ \forall i\}$ est stable par intersections finies et elle satisfait la relation $\sigma(\mathcal{C}) = \mathcal{E}_1 \times \dots \times \mathcal{E}_n$. Donc par le Corollaire 8, on conclut que $X_1, \dots, X_n$ sont indépendant si et seulement si:

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

Dans ce cas on obtient par Fubini-Tonelli:

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] &= \int \prod_{i=1}^n f_i(x_i)\, \mathbb{P}_{(X_1, \dots, X_n)}(dx_1 \dots dx_n) = \int \prod_{i=1}^n f_i(x_i)\, \mathbb{P}_{X_1}(dx_1) \cdots \mathbb{P}_{X_n}(dx_n) \\ &= \prod_{i=1}^n \int f_i(x_i)\, \mathbb{P}_{X_i}(dx_i) = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]\end{aligned}$$

$\square$

Note: La proposition 11 indique comment construire des VA indépendantes avec lois $\mu_1, \dots, \mu_n$ sur $(E_1, \mathcal{E}_1) \cdots (E_n, \mathcal{E}_n)$.

$$\Omega = E_1 \times \dots \times E_n,\ \mathcal{A} = \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n,\ \mathbb{P} = \mu_1 \otimes \dots \otimes \mu_n,\ X_i(\omega_1, \dots, \omega_n) := \omega_i \quad i = 1, \dots, n$$

Alors $\mathbb{P}_{(X_1, \dots, X_n)} = \mu_1 \otimes \dots \otimes \mu_n$

Remarque 12

(i) Si $X_1, \ldots, X_n$ sont indépendantes et $f_i : E_i \to \mathbb{R}$ mesurables, on a:

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

a condition que $\mathbb{E}[\,|f_i(X_i)|\,] < \infty \ \forall i$.
(En particulier si $X_1, \ldots, X_n \in L^1$ sont indépendantes alors $X_1 \cdots X_n \in L^1$. Faux en général ex: $X_1 = X_2$, $X_1, X_2 \in L^1$ et cela impliquerait que $X_1 \cdot X_1 = X_1^2 \in L^1 \implies X_1 \in L^2$ et donc pour tout X dans L^1, X est aussi L^2, ce qui est faux.

(ii) Si $X_1, X_2 \in L^2$ sont indépendantes alors $\mathrm{Cov}(X_1, X_2) = 0$
Attention: La réciproque est fausse. Exemple: X_1 une VA avec densité p symétrique, i.e. $p(-x) = p(x)$ et $\varepsilon \in \{\pm 1\}$ avec $\mathbb{P}(\varepsilon = +1) = \mathbb{P}(\varepsilon = -1) = \dfrac{1}{2}$

Alors X_1 et $X_2 := \varepsilon X_1$ satisfont $\mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1 \cdot X_2] - \underbrace{\mathbb{E}[X_1]}_{=0} \cdot \mathbb{E}[X_2] = \mathbb{E}[\varepsilon X_1^2] = \underbrace{\mathbb{E}[\varepsilon]}_{=0} \cdot \mathbb{E}[X_1^2] = 0$

Néanmoins, X_1 et X_2 ne sont pas indépendantes. Si elles l'étaient alors $|X_1|$ et $|X_2| = |X_1|$ seraient indépendantes. Or si VA réelle X est indépendante d'elle même est est presque sûrement constante.

⌜

Soit X une telle VA:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 0 \implies \forall t > 0,\ \mathbb{P}(\,|X - \mathbb{E}[X]| > t\,) \overset{\text{Chebychev}}{\leq} \frac{\mathrm{Var}(X)}{t^2} = 0 \implies X = \mathbb{E}[X] \quad \text{p.s.}$$

⌟

Mais X_1 n'est pas p.s. constante vu qu'elle a la densité $2 \cdot p(x) \cdot 1_{\mathbb{R}_+(x)}$

(iii) Remarquablement, si la loi de (X_1, X_2) est gaussienne, alors on a:

$$\mathrm{Cov}(X_1, X_2) = 0 \iff X_1 \quad \text{et} \quad X_2 \quad \text{indépendantes}$$

(Théorème de Wick, exo. 4.2)

Remarque 13

Soient X, Y, Z des VA indépendantes. Alors X et $f(Y, Z)$ sont indépendants pour toute fonction mesurable f.

⌜

$$\begin{aligned}\mathbb{P}(X \in A, f(Y,Z) \in B) &= \mathbb{P}\big(\,(X,Y,Z) \in A \times f^{-1}(B)\,\big) = (\mathbb{P}_X \otimes \mathbb{P}_Y \otimes \mathbb{P}_Z)\big(A \times f^{-1}(B)\big) \\ &= \mathbb{P}_X(A) \cdot (\mathbb{P}_Y \otimes \mathbb{P}_Z)\big(f^{-1}(B)\big) = \mathbb{P}(X \in A) \cdot \mathbb{P}(\,f(Y,Z) \in B\,)\end{aligned}$$

⌟

Ce principe de regroupement par paquets se généralise facilement.

2.4 Le lemme de Borel-Cantelli

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'évènements. On définit:

$$\limsup_n A_n := \bigcap_{n\geq 0}\bigcup_{k\geq n} A_k \quad \text{et} \quad \liminf_n A_n := \bigcup_{n\geq 0}\bigcap_{k\geq n} A_k$$

Intérprétation: $\limsup_n A_n \to$ Si A_k représente l'événement qu'il pleut au jour k, alors l'événement $\bigcap\bigcup A_k$ représente l'événement qu'il pleut une infinité de jours.
En reprenant cette interprétation $\bigcup\bigcap A_k$ représenterait le fait qu'il existe un jour k à partir duquel il pleut tout les jours. Plus mathématiquement:

$$\limsup_n A_n = \{\omega \in A_n \quad \text{un nombre infini de fois} \quad \}$$
$$\liminf_n A_n = \{\omega \in A_n \quad \text{finalement} \quad \}$$

On a $\liminf_n A_n \subset \limsup_n A_n$.

Proposition 14 (Lemme de Borel-Cantelli)

(i) Si $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}\left(\limsup_n A_n\right) = 0$, i.e.: $\{n \in \mathbb{N} \mid \omega \in A_n\}$ est fini p.s.

(ii) Si $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ et si $(A_n)_{n\in\mathbb{N}}$ sont indépendants alors $\mathbb{P}\left(\limsup_n A_n\right) = 1$, i.e.: $\{n \in \mathbb{N} \mid \omega \in A_n\}$ est infini p.s.

Remarque: Dans (ii), l'indépendance est importante; considérer:

$$A_n = A \quad \forall n \in \mathbb{N} \quad \text{avec} \quad 0 < \mathbb{P}(A) < 1$$

<u>*Démonstration:*</u>

(i)

$$\mathbb{P}\left(\limsup_n A_n\right) = \lim_n \mathbb{P}\left(\bigcup_{k\geq n} A_k\right) \leq \lim_n \sum_{k\geq n} \mathbb{P}(A_k) \overset{\sum \mathbb{P}(A_n)<\infty}{=} 0$$

(ii) Pour un $N \in \mathbb{N}$ fixé et $n \geq N$ on a:

$$\mathbb{P}\left(\bigcap_{k=N}^{n} A_k^c\right) = \prod_{k=N}^{n} \mathbb{P}(A_k^c) = \prod_{k=N}^{n}\big(1-\mathbb{P}(A_k)\big) \overset{1-x\leq e^{-x}}{\leq} \prod_{k=N}^{n} \exp\big(-\mathbb{P}(A_k)\big) = \exp\left(-\sum_{k=N}^{n} \mathbb{P}(A_k)\right) \overset{n\to\infty}{\longrightarrow} 0$$

Donc $\mathbb{P}\left(\bigcap_{k\geq N} A_k^c\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=N}^{n} A_k^c\right) = 0 \implies \mathbb{P}\left(\bigcup_{N\geq 0}\bigcap_{k\geq N} A_k^c\right) = 0 \overset{\text{de Morgan}}{\implies} \mathbb{P}\left(\bigcap_{N\geq 0}\bigcup_{k\geq N} A_k\right) = 1.$

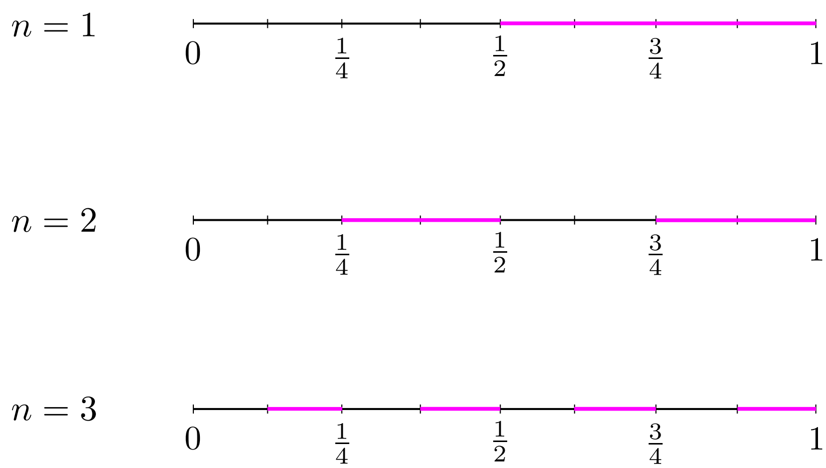
$\square$

Exemple 15

Soit $(\Omega, \mathcal{A}, \mathbb{P}) = (\ [0,1),\ \mathcal{B}\big([0,1)\big),\ \lambda\ )$ (λ mesure de Lebesgue). Pour $n \geq 1$ on définit l'évènement:

$$B_n := \bigcup_{k=1}^{2^{n-1}} \left[\frac{k-\frac{1}{2}}{2^{n-1}},\ \frac{k}{2^{n-1}}\right)$$

Nous pouvons représenter les 3 premiers B_n comme ceci:



On définit ensuite $X_n := 1_{B_n}$. Si $\omega = \sum_{k \geq 1} x_k \cdot 2^{-k}$ est la représentation en binaire de ω, on a pour tout $n \geq 1$:

$$\omega = \sum_{k=1}^{n} x_k \cdot 2^{-k} + \varepsilon_n \quad \text{où} \quad 0 \leq \varepsilon < \frac{1}{2^n}$$

et on vérifie par récurrence que $X_n(\omega) = x_n$. Donc $X_n(\omega) = n$-ème chiffre après la virgule de ω (en binaire).

De plus, $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = \dfrac{1}{2}$ et $(X_n)_{n \geq 1}$ est une famille indépendante.

En effet, pour tout $x_1, \ldots, x_p \in \{0,1\}$ on a:

$$\mathbb{P}(X_1 = x_1, \ldots, X_p = x_p) = \mathbb{P}\left(\left[\sum_{k=1}^{p} x_n \cdot 2^{-k},\ \sum_{k=1}^{p} x_k \cdot 2^{-k} + 2^{-p}\right)\right) = \frac{1}{2^p} = \prod_{k=1}^{p} \mathbb{P}(X_k = x_k)$$

Soit $p \in \mathbb{N}^*$ et $x_1, \ldots, x_p \in \{0,1\}$. Alors on a p.s.:

$$\#\{k \mid (X_{k+1}, \ldots, X_{k+p}) = (x_1, \ldots, x_p)\ \} = \infty$$

Pour $n \in \mathbb{N}$ posons:

$$Y_n := (X_{np+1}, \ldots, X_{np+p}) \quad \rightsquigarrow \quad \underbrace{x_1, \ldots, x_p}_{Y_0},\ \underbrace{x_{p+1}, \ldots, x_{2p}}_{Y_1},\ \underbrace{x_{2p+1}, \ldots, x_{3p}}_{Y_2}, \ldots$$

Par la remarque 13, $(Y_n)_{n \in \mathbb{N}}$ est une famille indépendante. Posons $A_n := \{Y_n = (x_1, \ldots, x_p)\}$.
Alors (A_n) est indépendant et $\mathbb{P}(A_n) = 2^{-p}$. Appliquer 14(ii).

Plus fort (par intersection dénombrable):
p.s. $\forall p \geq 1$, $\forall x_1, \ldots, x_p \in \{0, 1\}$

$$\#\{k \in \mathbb{N}^* \mid (X_{k+1}, \ldots, X_{k+p}) = (x_1, \ldots, x_p)\ \} = \infty$$

Conclusion: Pour presque tout réel, n'importe quelle suite finie de 0 et 1 apparait une infinité de fois dans son développement binaire. Il est possible de montrer le même résultat pour le développement décimal d'un nombre.

2.5 Sommes de variables aléatoires indépendantes

Définition 16

Soient μ et ν deux mesures de probabilité sur $\mathbb{R}$. La <u>convolution</u> de μ et ν est la mesure:

$$\mu * \nu := p_*(\mu \otimes \nu) \quad \text{où} \quad p(x,y) := x + y$$

Remarque 17

Si $\mu(dx) = f(x)\,dx$ et $\nu(dx) = g(x)\,dx$ alors:

$$(\mu * \nu)(dx) = h(x)\,dx \quad \text{où} \quad h(x) = \int f(x-y)g(y)\,dy \quad \big(\equiv \big(f * g\big)(x)\big)$$

⌜

Pour $\varphi \geq 0$ on a:

$$\int \varphi(x)\,(\mu * \nu)(dx) = \int \varphi(x+y)\,(\mu \otimes \nu)(dx\,dy) = \int \varphi(x+y) f(x) g(y)\,dx\,dy = \int \varphi(x) \left(\int f(x-y)g(y)\,dy \right) dx$$

⌟

Remarque 18

Si X et Y sont indépendantes alors:

$$\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y \quad \text{De plus:} \quad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Notation: $X \overset{\mathrm{d}}{=} Y \iff \mathbb{P}_X = \mathbb{P}_Y$

Exemple 19

Soit $X_1 \overset{\mathrm{d}}{=} \mathcal{N}(m_1, a_1)$ et $X_2 \overset{\mathrm{d}}{=} \mathcal{N}(m_2, a_2)$ indépendantes. Alors:

$$X_1 + X_2 \overset{\mathrm{d}}{=} \mathcal{N}(m_1 + m_2, a_1 + a_2) \quad \text{Exo.}$$

Proposition 20: Loi faible des grands nombres

Soient $(X_n)_{n \in \mathbb{N}^*}$ des VA indépendantes dans L^2 avec la même loi. Alors:

$$\frac{1}{n}(X_1 + \cdots + X_n) \quad \xrightarrow[n\to\infty]{\in L^2} \quad \mathbb{E}[X_1]$$

<u>*Démonstration:*</u>

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}(X_1 + \cdots + X_n) - \mathbb{E}[X_1] \right)^2 \right] &= \mathbb{E}\left[\left(\frac{1}{n}(X_1 + \cdots + X_n) - \mathbb{E}\left[\frac{1}{n}(X_1 + \cdots + X_n) \right] \right)^2 \right] \\
&= \mathrm{Var}\left(\frac{1}{n}(X_1 + \cdots + X_n) \right) = \frac{1}{n^2} \mathrm{Var}(X_1 + \cdots + X_n) \\
&= \frac{1}{n^2} \left(\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) \right) \\
&= \frac{1}{n} \mathrm{Var}(X_1)
\end{aligned}$$

□

Remarque: Il suffirait que $\mathrm{Cov}(X_i, X_j) = 0 \ \forall i \neq j$ à la place de l'indépendance. En effet reprenons l'égalité de la démonstration:

$$\frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n) = \frac{1}{n^2}\left(\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)\right)$$

On utilise, avec $Y_n = X_n - \mathbb{E}[X_n]$:

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathbb{E}\left[\left(\sum_{i=1}^n Y_i\right)^2\right] = \sum_{i,j=1}^n \underbrace{\mathbb{E}[Y_iY_j]}_{=0 \text{ si } i\neq j} = \sum_{i=1}^n \mathbb{E}[Y_i^2] = \sum_{i=1}^n \mathrm{Var}(X_i)$$

Exemple 21: Approximation polynômiale

Soit f continue sur $[0,1]$ et définissons le n-ème polynôme de Bernstein:

$$f_n(x) := \sum_{k=0}^n \binom{n}{k} \cdot x^k \cdot (1-x)^{n-k} \cdot f\left(\frac{k}{n}\right)$$

Alors $f_n \to f$ uniformément.

⌜

Soient $X_1, \ldots, X_n$ indépendantes avec la loi Bernoulli (p). Alors $S_n = X_1 + \cdots + X_n$ a la loi binômiale:

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{et notons aussi} \quad \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = f_n(p)$$

Intuitivement, on s'attend à ce que $\mathbb{E}\big[f(S_n/n)\big] \to f(p)$ vu que $\dfrac{S_n}{n} \xrightarrow{\in L^2} p$.
Plus précisément: soit $\varepsilon > 0$ et choisissons $\delta > 0$ tel que $|x-y| < \delta \implies |f(x) - f(y)| < \varepsilon$ (f est continue sur un compact et donc elle est aussi uniformément continue.) Alors:

$$\begin{aligned}
\left|\mathbb{E}\big[f\left(\frac{S_n}{n}\right)\big] - f(p)\right| &\leq \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(p)\right|\right] \cdot \left(1_{\left|\frac{S_n}{n}-p\right|<\delta} + 1_{\left|\frac{S_n}{n}-p\right|\geq\delta}\right) \\
&\leq \varepsilon + 2\|f\|_{L^\infty} \cdot \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right) \\
\text{Par Chebychev:} \quad &\leq \varepsilon + 2\|f\|_{L^\infty} \cdot \mathrm{Var}\left(\frac{S_n}{n}\right) \cdot \frac{1}{\delta^2} \leq \varepsilon + 2\|f\|_{L^\infty} \cdot \frac{p\,(1-p)}{n\cdot\delta^2} \underset{n\to\infty}{\longrightarrow} \varepsilon
\end{aligned}$$

⌟

Dans Prop. 20, "faible" veut dire dans L^2 plutôt que p.s. Une convergence p.s. est beaucoup plus utile en général.

Remarque 22

Si $Y_n \to Y$ dans L^p pour $p \geq 1$, il se pourrait très bien que $Y_n(\omega) \not\to Y(\omega)$ pour tout $\omega \in \Omega$

Pour un exemple, $a_0 = 0, a_1, a_2, \cdots \in [0,1)$ avec $a_n \overset{n\to\infty}{\longrightarrow} 0$, $\displaystyle\sum_{n\in\mathbb{N}} a_n = \infty$

Soit $B_n := \left[\displaystyle\sum_{k=0}^n a_k, \sum_{k=0}^{n+1} a_k\right)$. On peut ensuite définir $A_n := \phi(B_n)$ où $\phi(x) := x - \lfloor x \rfloor$ (ϕ(x) rend la partie décimale de x)
Soit $\Omega = [0,1)$, $\mathbb{P} =$ Lebesgue. Alors: $\mathbb{P}(A_n) = \lambda(B_n) = a_n$. Soit $Y_n := 1_{A_n}$, alors:

$$\mathbb{E}[\,|Y_n|^p\,] = a_n \underset{n\to\infty}{\longrightarrow} 0 \quad \forall p \geq 1$$

Mais $\forall \omega \in [0,1)$ on a $Y_n(\omega) = 1$ pour un nombre infini de n.
Note: Si $\displaystyle\sum_n a_n < \infty$ alors $Y_n \to 0$ p.s.

Proposition 23: Loi des grands nombres dans L^4

Soient $(X_n)_{n\geq 1}$ des VA dans L^4 avec la même loi. Alors p.s.

$$\frac{1}{n}(X_1+\cdots+X_n)\to\mathbb{E}[X_1]$$

<u>*Démonstration:*</u>
On peut remplacer X_n par $X_n-\mathbb{E}[X_n]$ et donc supposer que les VA ont des éspérences nulles ($\mathbb{E}[X_n]=0$). Alors:

$$\mathbb{E}\left[\big(\frac{1}{n}(X_1+\cdots+X_n)\big)^4\right]=\frac{1}{n^4}\sum_{k_1,k_2,k_3,k_4=1}^{n}\mathbb{E}\big[X_{k_1}X_{k_2}X_{k_3}X_{k_4}\big]$$

On voit ici que les termes (k_1,k_2,k_3,k_4) doivent être regroupés 2 par 2 (ou par 4), d'où la séparation de somme:

$$\begin{aligned}&=\frac{1}{n^4}\left(n\cdot\mathbb{E}[\ X_1^4\ ])+3\cdot n(n-1)\mathbb{E}[\ X_1^2\ ]\cdot\mathbb{E}[\ X_1^2\ ]\right)\\&\leq\frac{C}{n^2}\quad\text{Pour une constante}\quad C<\infty\end{aligned}$$

On déduit:

$$\sum_{n\geq 1}\mathbb{E}\left[\ \big(\frac{1}{n}(X_1+\cdots+X_n)\big)^4\ \right]=\mathbb{E}\left[\sum_{n\geq 1}\big(\frac{1}{n}(X_1+\cdots+X_n)\big)^4\right]<\infty$$

Et donc

$$\sum_{n\geq 1}\big(\frac{1}{n}(X_1+\cdots+X_n)\big)^4<\infty\quad\text{p.s.}$$

$\square$

Corollaire 24

$(A_n)_{n\in\mathbb{N}^*}$ événements indépendants de même probabilité.

$$\implies\ \frac{1}{n}\sum_{i=1}^{n}1_{A_i}\to\mathbb{P}(A_1)$$

Exemple 15 (suite)

Soit $p\in\mathbb{N}^*$ et $\ell\in\{1,\ldots,p\}$, et définissons

$$Y_n^\ell:=\big(X_{(n-1)p+\ell}\ ,\ldots,X_{np+\ell-1}\big)$$

Par l'exemple 15 et la remarque 13, $\forall\ell\in\{1,\ldots,p\}$, $(Y_n^\ell)_{n\in\mathbb{N}^*}$ est une famille indépendante. Pour tout $x=(x_1,\ldots,x_p)\in\{0,1\}^p$, en appliquant le corollaire 24 à l'événement $A_n=\{Y_n^\ell=x\}$ on obtient:

$$\frac{1}{n}\#\big\{i\in\{1,\ldots,n\}\mid Y_n^\ell=x\big\}\xrightarrow[n\to\infty]{\text{p.s.}}2^{-p}$$

Vu que ceci est vrai pour tout $\ell\in\{1,\ldots,p\}$ on trouve:

$$\frac{1}{n}\#\big\{i\in\{1,\ldots,n\}\mid\big(X_i,\ldots,X_{i+p-1}\big)=x\big\}\xrightarrow[n\to\infty]{\text{p.s.}}2^{-p}$$

En prenant l'intersection (dénombrable) sur $p\in\mathbb{N}^*$ et $x\in\{0,1\}^p$ on trouve: Pour presque tout $\omega\in[0,1)$, on a $\forall p\in\mathbb{N}^*$, $\forall x\in\{0,1\}^p$,

$$\frac{1}{n}\#\big\{i\in\{1,\ldots,n\}\mid\big(X_i(\omega),\ldots,X_{i+p-1}(\omega)\big)=x\big\}\xrightarrow[n\to\infty]{}2^{-p}$$

3 Convergence de variables aléatoires

3.1 Notions de convergence

Soient $(X_n)_{n\geq 1}$, X des VA à valeurs dans $\mathbb{R}$. Que veut dire $X_n \underset{n\to\infty}{\longrightarrow} X$?

Rappel:

- $X_n \xrightarrow[n\to\infty]{\text{p.s.}} X$ si $\mathbb{P}(\lim_n X_n = X) = 1$
- $X_n \xrightarrow[n\to\infty]{L^p} X$ pour $p \geq 1$ si $\lim_n \mathbb{E}\left[\, \left|X_n - X\right|^p \,\right] = 0$.

Définition 1

X_n converge en probabilité vers X, noté $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, si $\forall \varepsilon > 0, \quad \lim_n \mathbb{P}\left(\, \left|X_n - X\right| > \varepsilon\right) = 0$

Proposition 2

Soit $\mathcal{L}^0$ l'espace des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$ et $L^0 = \mathcal{L}^0 /\sim$, où $X \sim Y \iff X = Y$ p.s. Posons les opérateurs: $x \wedge y := \min(x, y)$ et $x \vee y := \max(x, y)$ et définissons:

$$d(X, Y) := \mathbb{E}\big[\, |X - Y| \wedge 1 \,\big]$$

Et alors (L^0, d) est un espace métrique complet, et on a $X_n \xrightarrow{\mathbb{P}} X \iff d(X_n, X) \to 0$

Démonstration:

- d est une métrique sur L^0: évident.
- Supposons que $X_n \xrightarrow{\mathbb{P}} X \implies \forall \varepsilon \in (0, 1]$:

$$\begin{aligned}\mathbb{E}\big[\, |X_n - X| \wedge 1 \,\big] &= \mathbb{E}\big[\, |X_n - X| \cdot 1_{|X_n - X| \leq \varepsilon}\big] + \mathbb{E}\big[\, |X_n - X| \cdot 1_{|X_n - X| > \varepsilon}\big] \\ &\leq \varepsilon + \mathbb{P}(\, |X_n - X| > \varepsilon \,) \to \varepsilon \implies d(X_n, X) \to 0\end{aligned}$$

- $d(X_n, X) \to 0 \implies \forall \varepsilon \in (0, 1]$:

$$\mathbb{P}\,(\, |X_n - X| > \varepsilon \,) \overset{\substack{\text{Cheby.} \\ f(x) = x \wedge 1}}{\leq} \frac{\mathbb{E}\big[\, |X_n - X| \,\big]}{\varepsilon} \to 0 \implies X_n \xrightarrow{\mathbb{P}} X$$

- Reste à montrer que L^0 est complet. Soit $\left(X_n\right)$ une suite de Cauchy dans (L^0, d). Prenons une sous-suite $Y_k = X_{n_k}$ telle que $d(Y_k, Y_{k+1}) \leq 2^{-k}$. Alors:

$$\mathbb{E}\left[\, \sum_{k=1}^{\infty} \big(\, |Y_{k+1} - Y_k| \wedge 1\big) \,\right] \leq \sum_{k=1}^{\infty} 2^{-k} < \infty$$

$$\implies \sum_{k=1}^{\infty} (\, |Y_{k+1} - Y_k| \wedge 1) < \infty \quad \text{p.s.} \implies \sum_{k=1}^{\infty} |Y_{k+1} - Y_k| < \infty \quad \text{p.s.}$$

On peut alors définir:

$$X := Y_1 + \sum_{k=1}^{\infty} \big(Y_{k+1} - Y_k\big)$$

Alors $Y_k \xrightarrow{\text{p.s.}} X$ et donc:

$$d(Y_k, X) = \mathbb{E}\big[\, |Y_k - X| \wedge 1 \,\big] \underset{k\to\infty}{\overset{\substack{\text{conv.}\\ \text{dominée}}}{\longrightarrow}} 0$$

Donc $d(X_n, X) \to 0$ aussi.

$\square$

Proposition 3

(i) $X_n \xrightarrow{\text{p.s.}} X$ ou $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathbb{P}} X$

(ii) $X_n \xrightarrow{\mathbb{P}} X \implies \exists$ sous-suite (X_{n_k}) telle que $X_{n_k} \underset{k\to\infty}{\xrightarrow{\text{p.s.}}} X$

<u>*Démonstration:*</u>

(ii) Démonstartion de Prop. 2.

(i) $X_n \xrightarrow{\text{p.s.}} X \implies \mathbb{P}\big(\, |X_n - X| > \varepsilon \,\big) = \mathbb{E}\big[\, 1_{|X_n - X| > \varepsilon} \,\big] \overset{\substack{\text{conv.}\\ \text{dominée}}}{\longrightarrow} 0$

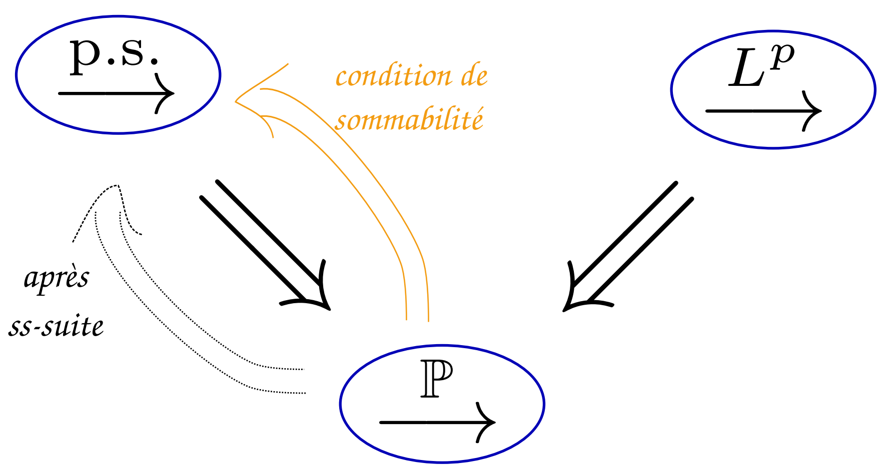
$X_n \xrightarrow{L^p} X \implies \mathbb{P}\big(\, |X_n - X| > \varepsilon \,\big) \leq \dfrac{\mathbb{E}\big[\, |X_n - X|^p \,\big]}{\varepsilon^p} \to 0$

$\square$

Remarque: Dans la Prop. 3 (ii), prendre une sous-suite est en général nécéssaire. (cf. Remarque 2.22) Mais si la suite converge probabilistiquement suffisement vite: $\forall \varepsilon > 0 \quad \sum_{n\geq 1} \mathbb{P}(\, |X_n - X| > \varepsilon \,) < \infty$ alors il n'y a pas besoin de prendre une ss-suite, et $X_n \xrightarrow{\text{p.s.}} X$, par Borel-Cantelli:

$$\mathbb{E}\left[\sum_{n\geq 1} 1_{|X_n - X| > \varepsilon} \right] < \infty$$

Résumé des notions de convergence de VA sur $(\Omega, \mathcal{A}, \mathbb{P})$



3.2 Loi des grands nombres

But: Montrer la Proposition 2.23 sans l'hypothèse $X_1 \in L^4$.

Lemme 4

Pour $i = 1, \dots, n$ soit $\mathcal{C}_i$ une famille stable par intersections finies contenant Ω, et soit $\mathcal{B}_i := \sigma(\mathcal{C}_i)$. Si $\forall C_1 \in \mathcal{C}_1, \dots, C_n \in \mathcal{C}_n$:

$$\mathbb{P}\left(\bigcap_i C_i\right) = \prod_i \mathbb{P}(C_i)$$

alors $\mathcal{B}_1, \dots, \mathcal{B}_n$ sont indépendants.
Démonstration:
Pour $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$ fixés, on pose:

$$\mathcal{M}_1 := \left\{B_1 \in \mathcal{B}_1 \mid \mathbb{P}\big(B_1 \cap C_2 \cap \cdots \cap C_n\big) = \mathbb{P}(B_1) \cdot \mathbb{P}(C_2) \cdots \mathbb{P}(C_n)\right\}$$

Alors $\mathcal{C}_1 \subset \mathcal{M}_1$ et $\mathcal{M}_1$ est une classe monotone, donc par le lemme de classe monotone.

$$\mathcal{M}_1 \supset \mathcal{M}(\mathcal{C}_1) = \sigma(\mathcal{C}_1) = \mathcal{B}_1$$

et on a donc $\forall B_1 \in \mathcal{B}_1, C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$: $\mathbb{P}\big(B_1 \cap C_2 \cap \cdots \cap C_n\big) = \mathbb{P}(B_1)\mathbb{P}(C_2) \cdots \mathbb{P}(C_n)$
On continue ainsi: fixons $B_1 \in \mathcal{B}_1, C_3 \in \mathcal{C}_3, \dots, C_n \in \mathcal{C}_n$ et définissons:

$$\mathcal{M}_2 := \left\{B_2 \in \mathcal{B}_2 \mid \mathbb{P}\big(B_1 \cap B_2 \cap C_3 \cap \cdots \cap C_n\big) = \mathbb{P}(B_1) \cdot \mathbb{P}(B_2) \cdot \mathbb{P}(C_3) \cdots \mathbb{P}(C_n)\right\}$$

Par le lemme de classe monotone et la conclusion du l'étape précédente, on a $\mathcal{M}_2 \supset \mathcal{B}_2$... continuer par récurrence.

$\square$

Proposition 5: Loi du tout ou rien, loi de 0-1

Soient $(X_n)_{n\geq 1}$ des VA indépendantes. Pour $n \geq 1$ on définit la tribu: $\mathcal{B}_n := \sigma\left(X_k \mid k \geq n\right)$. Alors la tribu asymptotique:

$$\mathcal{B}_\infty := \bigcap_{n \geq 1} \mathcal{B}_n$$

est grossière, au sens où $\mathbb{P}(B) = 0$ ou $\mathbb{P}(B) = 1$, $\forall B \in \mathcal{B}_\infty$

Démonstration:
Soit $\mathcal{D}_n := \sigma(X_k \mid k \leq n)$. Alors $\mathcal{D}_n$ et $\mathcal{B}_{n+1}$ sont indépendantes.

⌜

(i) Pour $k \geq n+1$ soit $\mathcal{B}_{n+1,k} = \sigma(X_{n+1}, \dots, X_k)$.
En prenant $\mathcal{C}_1 = \{B_1 \cap \cdots \cap B_n \mid B_i \in \sigma(X_i)\}$ et $\mathcal{C}_2 = \{B_{n+1} \cap \cdots \cap B_k \mid B_i \in \sigma(X_i)\}$
Dans le Lemme 4, on trouve que $\mathcal{D}_n$ et $\mathcal{B}_{n+1,k}$ sont indépendants.

(ii) Soit $\mathcal{C}_1 = \mathcal{D}_n$ et $\mathcal{C}_2 = \bigcup_{k \geq n+1} \mathcal{B}_{n+1,k}$. Alors $\sigma(\mathcal{C}_1) = \mathcal{D}_n$ et $\sigma(\mathcal{C}_2) = \mathcal{B}_{n+1}$, et par (i) et Lemme 4 nous avons que $\mathcal{D}_n$ et $\mathcal{B}_{n+1}$ sont indépendantes.

⌟

Donc $\mathcal{D}_n$ et $\mathcal{D}_\infty$ sont indépendants $\forall n$.
Avec $\mathcal{C}_1 = \bigcup\limits_{n \geq 1} \mathcal{D}_n$ et $\mathcal{C}_2 = \mathcal{B}_\infty$ on a donc $\forall C_1 \in \mathcal{C}_1,\ C_2 \in \mathcal{C}_2$:

$$\mathbb{P}(C_1 \cap C_2) = \mathbb{P}(C_1) \cdot \mathbb{P}(C_2)$$

donc par le Lemme 4, cela implique:

$$\sigma(\mathcal{C}_1) = \sigma(x_n \mid n \geq 1) \quad \text{et} \quad \sigma(\mathcal{C}_2) = \mathcal{B}_\infty \quad \text{sont indépendants}$$

Puisque $\mathcal{B}_\infty \subset \sigma(x_n \mid n \geq 1)$ on a que $\mathcal{B}_\infty$ est indépendante d'elle même. $\forall B \in \mathcal{B}_\infty$:

$$\mathbb{P}(B) = \mathbb{P}(B \cap B) = \mathbb{P}(B)^2 \implies \mathbb{P}(B) \in \{0, 1\}$$

$\square$

Remarque 6

Si X est mesurable par rapport à une tribu grossière, alors $X = \text{const.}$ presque sûrement.

Soit $(X_n)_{n \geq 1}$ des VA indépendantes. Alors:

$$X := \limsup_{k \to \infty} \frac{1}{k}\,(X_1 + \cdots + X_k) = \limsup_{k \to \infty} \frac{1}{k}\,(X_n + \cdots + X_k)$$

Et ce $\forall n \geq 1$, donc X st $\mathcal{B}_n$-mesurable $\forall n \geq 1$, donc X est $\mathcal{B}$-mesurable. Même chose pour $\liminf$.

Conclusion: si $\dfrac{1}{n}(X_1 + \cdots + X_n)$ converge p.s. alors la limite est constante p.s.
Si les X_n ont la même loi, par Exercice 8.3, on sait que si $X_1 \notin L^1$ alors $\dfrac{1}{n}(X_1 + \cdots + X_n)$ diverge p.s.

Proposition 7: Loi des grands nombres

Soient $(X_n)_{n \geq 1}$ des VA indépendantes, de même loi, dans L^1. Alors:

$$\frac{1}{n}(X_1 + \cdots + X_n) \quad \xrightarrow[n \to \infty]{\text{p.s.}} \quad \mathbb{E}\big[X_1\big]$$

<u>*Démonstration:*</u>
Soient $S_n := X_1 + \cdots X_n$ et $S_0 := 0$. Soit $a > \mathbb{E}\big[X_1\big]$ et $M := \sup\limits_{n \in \mathbb{N}}(S_n - n \cdot a)$
M est une VA à valeurs dans $[0, \infty]$ (pour $n = 0$ on a $M = 0$ donc le sup est forcément au dessus)
On va montrer que:

$$M < \infty \quad \text{p.s.} \qquad (\star)$$

Supposant que $(\star)$ soit montré, de $S_n \leq na + M \quad \forall n$ on déduit:

$$L := \limsup_{n \to \infty} \quad \frac{S_n}{n} \leq a \quad \text{p.s.} \quad \forall a > \mathbb{E}\big[X_1\big]$$

Donc $L \leq \mathbb{E}\big[X_1\big]$ p.s.

$$\forall k \in \mathbb{N}^* \quad \text{on a} \quad \mathbb{P}\big(L \leq \overbrace{\mathbb{E}\big[X_1\big] + \frac{1}{k}}^{a}\big) = 1 \implies \mathbb{P}\Big(\bigcap_{k\in\mathbb{N}^*} \underbrace{\{L \leq \mathbb{E}\big[X_1\big] + \frac{1}{k}\}}_{\{L\leq\mathbb{E}[X_1]\}}\Big) = 1$$

En remplaçant X_n par $-X_n$ on obtient que:

$$\liminf_{n\to\infty} \quad \frac{S_n}{n} \geq \mathbb{E}\big[X_1\big] \quad \text{p.s.}$$

On conclut: $\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}\big[X_1\big]$ p.s.

Il reste à démontrer $(\star)$.
D'abord, $\{M < \infty\} \in \mathcal{B}_\infty$

$$\begin{aligned}\{M < \infty\} &= \Big\{\sup_{n\geq 0}(S_n - na) < \infty\Big\} \\ &= \Big\{\sup_{n\geq k}\big(S_n - S_k - (n-k)a\big) < \infty\Big\} \in \mathcal{B}_{k+1} = \sigma\big(X_{k+1}, X_{k+2}, \dots\big), \quad \forall k \geq 0\end{aligned}$$

Par la loi du tout ou rien, il suffit de montrer $\mathbb{P}(M = \infty) < 1$. On procèdera par l'absurde en supposant que $M = \infty$ p.s.

On définit $\forall k \in \mathbb{N}$

$$\begin{aligned}M_k &:= \sup_{0\leq n\leq k} \big(\underbrace{S_n}_{X_1+\cdots+X_n} - \; na\big) \\ M'_k &:= \sup_{0\leq n\leq k} \big(\underbrace{S_{n+1} - S_1}_{X_2+\cdots+X_{n+1}} - \; na\big)\end{aligned}$$

Clairement, $M_k \stackrel{\mathrm{d}}{=} M'_k \quad \forall k \in \mathbb{N}$. On a $M = \lim_{k\to\infty} M_k$ et on définit $M' := \lim_{k\to\infty} M'_k$. (Notons que $\{M_k\}_{k\in\mathbb{N}}$ et $\{M'_k\}_{k\in\mathbb{N}}$ sont des suites croissantes, ce qui justifie l'existance de leur limite.)

On conclut que $M \stackrel{\mathrm{d}}{=} M'$.

$$\mathbb{P}(M \leq x) = \mathbb{P}(M_k \leq x \quad \forall k) \stackrel{(*)}{=} \lim_{k\to\infty} \mathbb{P}(M_k \leq x)$$

$(*)$: Car M_k est croissante. On peut utiliser le même argument pour M'_k.

De plus:

$$M_{k+1} = \sup\big(0, \sup_{1\leq n\leq k+1}(S_n - na)\big)$$

En posant le changement de variable: $n \mapsto n+1$

$$\begin{aligned} &= \sup\big(0, \sup_{0\leq n\leq k}(S_{n+1} - (n+1)a)\big) = \sup\big(0,\ M'_k + X_1 - a\big) \\ &= M'_k - \inf\big(a - X_1,\ M'_k\big) \end{aligned}$$

Donc en prenant l'espérence des deux côté (tout fonctionne car nous sommes dans L^1):

$$\begin{aligned} \mathbb{E}\Big[\inf\big(a - X_1,\ M'_k\big)\Big] &= \underbrace{\mathbb{E}\big[M'_k\big]}_{=\mathbb{E}[M_k]} - \mathbb{E}\big[M_{k+1}\big] \\ &\leq 0 \end{aligned}$$

En utilisant que $\Big|\inf\big(a - X_1,\ \underbrace{M'_k}_{\geq 0}\big)\Big| \leq |a - X_1|$ et on peut alors utiliser la convergence dominée, on obtient:

$$\mathbb{E}\big[\inf\big(a - X_1,\ M'\big)\big] \leq 0$$

Si $M = \infty$ p.s. alors $M' = \infty$ p.s. car ils ont la même distribution, et donc $\inf\big(a - X_1,\ M'\big) = a - X_1$.
Donc $\mathbb{E}\big[a - X_1\big] \leq 0$ mais $\mathbb{E}\big[a - X_1\big] > 0$. ↯

$\square$

3.3 La convergence en loi

On note $C_b \equiv C_b(\mathbb{R}^d)$ pour l'ensemble des fonctions bornées (avec la norme sup) et continues sur $\mathbb{R}^d$.

Définition 8

(i) Soient (μ_n), μ des mesures de probobilité sur $\mathbb{R}^d$. Alors μ_n converge étroitement, ou faiblement, vers μ, noté $\mu_n \xrightarrow{e} \mu$, si:

$$\int \varphi \, d\mu_n \underset{n\to\infty}{\longrightarrow} \int \varphi \, d\mu \qquad \forall \varphi \in C_b$$

(Différentes notations existent dans la litérature: $\xrightarrow{d}$, $\xrightarrow{w}$, $\xrightarrow{w^*}$, $\Longrightarrow$)

(ii) Soient (X_n) et X des VA à valeurs dans $\mathbb{R}^d$. Alors X_n converge en loi, ou en distribution, noté $X_n \xrightarrow{d} X$, si $\mathbb{P}_{X_n} \xrightarrow{e} \mathbb{P}_X$. Ceci veut dire:

$$\mathbb{E}\big[\varphi(X_n)\big] \longrightarrow \mathbb{E}\big[\varphi(X)\big] \qquad \forall \varphi \in C_b$$

Remarque: La convergence $\xrightarrow{d}$ est de nature très différente de $\xrightarrow{\text{p.s.}}$, $\xrightarrow{L^p}$, $\xrightarrow{\mathbb{P}}$:
Elle ne concerne que les lois. En particulier, (X_n), X peuvent être définis sur des espaces de probabilité différents. De plus la limite n'est pas unique. Par exemple si X et Y ont la même loi, $X_n \xrightarrow{d} X$ et $X_n \xrightarrow{d} Y$.

Exemples 9

(i) Si $a_n \to a$ alors $\delta_{a_n} \xrightarrow{e} \delta_a$.

(ii) X_n est de loi uniforme sur $\left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\right\}$ et X de loi Lebesgue sur $[0,1]$. Alors $X_n \xrightarrow{d} X$.
(Approximation par sommes de Riemann pour fonctions continues.)

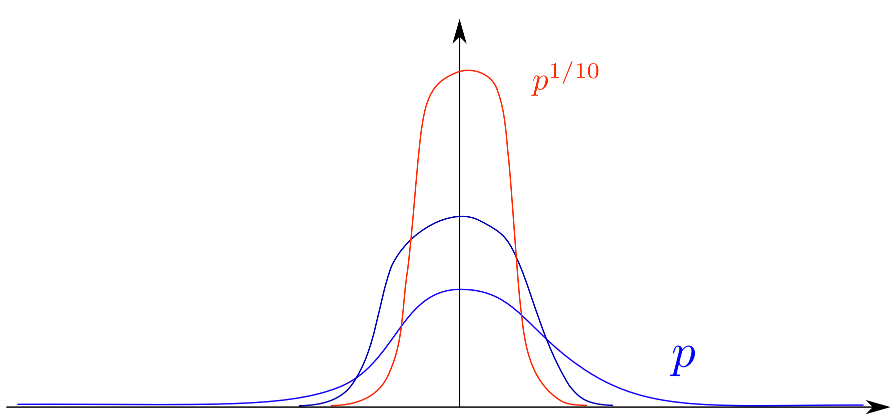
(iii) Soit μ une mesure de probabilité (sur $\mathbb{R}^d$) et $s^\eta(x) := \eta \cdot x$ pour $\eta > 0$ (une homothétie). Alors $s^\eta_* \mu \underset{\eta\to 0}{\xrightarrow{e}} \delta_0$

$\ulcorner$

$$\int \varphi(x) \; s^\eta_*\mu(dx) \overset{\text{Ex. 1.1}}{=} \int \varphi(\eta x) \; \mu(dx) \overset{\text{Conv. dominée}}{\longrightarrow} \varphi(0)$$

$\lrcorner$

En particulier, si $\mu(dx) = p(x)\, dx$, on a $s^\eta_*\mu(dx) = \frac{1}{\eta} \cdot p\left(\frac{x}{\eta}\right) \, dx$ (Ex. 2.2) En prenant p une gaussienne:



(iv) $\mu_n(dx) = p_n(x)$, $\mu(dx) = p(x)\,dx$. Si $\exists q \geq 0$ tel que $p_n \leq q \quad \forall n$ et $\int q(x)\,dx < \infty$ et $p_n \longrightarrow p$ presque partout, alors $\mu_n \xrightarrow{e} \mu$ par convergence dominée.

Proposition 10

Si $X_n \xrightarrow{\mathbb{P}} X$ alors $X_n \xrightarrow{d} X$.

Démonstration:

Supposons que $X_n \xrightarrow{\mathbb{P}} X$ mais que $X_n \not\xrightarrow{d} X$. Donc:

$$\exists \varphi \in C_b \quad \text{tel que} \quad \mathbb{E}\big[\varphi(X_n)\big] \underset{n\to\infty}{\not\longrightarrow} \mathbb{E}\big[\varphi(X)\big]$$

Prenons une sous-suite (n_k) et $\varepsilon > 0$ tel que $\forall k$: $\Big|\mathbb{E}\big[\varphi(X_{n_k})\big] - \mathbb{E}\big[\varphi(X)\big]\Big| \geq \varepsilon$.

Par Proposition 3(ii), il existe une sous-suite (n_{k_ℓ}) telle que $X_{n_{k_\ell}} \xrightarrow{\text{p.s.}} X$. Donc par convergence dominée on a:

$$\left|\mathbb{E}\left[\varphi\Big(X_{n_{k_\ell}}\Big)\right] - \mathbb{E}\big[\varphi(X)\big]\right| \underset{\ell\to\infty}{\longrightarrow} 0 \qquad \lightning$$

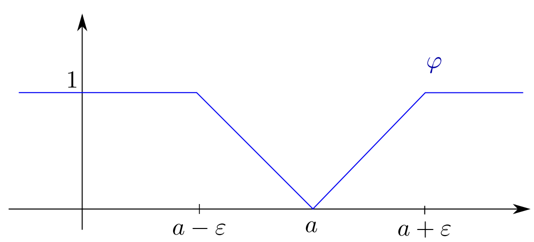
$\square$

Remarque 11

La réciproque est fausse. **Pire**: elle ne veut rien dire.
En général si $X_n \xrightarrow{d} X$ il se peut que (X_n) et X soient définis sur des espaces de probabilité différents.

Néanmoins, si $X_n \xrightarrow{d} a$, où a est une constante, alors $X_n \xrightarrow{\mathbb{P}} a$.

Soit $\varepsilon > 0$ et $\varphi \in C_b$ telle que $\varphi(a) = 0$ et $\varphi(x) = 1$ si $|x - a| \geq \varepsilon$ (par exemple $\varphi(x) = 1 \wedge \dfrac{|x-a|}{\varepsilon}$):



Alors:

$$\mathbb{P}\big(|X_n - X| > \varepsilon\big) = \mathbb{E}\left[1_{|X_n - X| > \varepsilon}\right] \leq \mathbb{E}\big[\varphi(X_n)\big] \underset{n\to\infty}{\longrightarrow} \varphi(a) = 0$$

Proposition 12

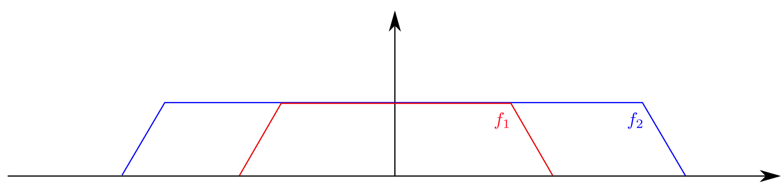
Soit $C_c = \{$ fonctions continues à support compacte $\}$ (nous sommes toujours dans $\mathbb{R}^d$) et $H \subset C_b$ tel que l'adhérence de H par $\|\cdot\|_{L^\infty}$ contient C_c . Sont équivalents:

(i)
$$\forall \varphi \in C_b,\ \int \varphi d\mu_n \longrightarrow \int \varphi\, d\mu \qquad \text{(c-à-d: } \mu_n \xrightarrow{e} \mu)$$

(ii)
$$\forall \varphi \in C_c,\ \int \varphi d\mu_n \longrightarrow \int \varphi\, d\mu$$

(iii)
$$\forall \varphi \in H,\ \int \varphi d\mu_n \longrightarrow \int \varphi\, d\mu$$

Démonstration:
(i) $\Longrightarrow$ (ii) et (i) $\Longrightarrow$ (iii): évidents.
On montre (ii) $\Longrightarrow$ (i) et (iii) $\Longrightarrow$ (ii). (ii) $\Longrightarrow$ (i):
Supposons (ii). Soit $\varphi \in C_b$. Soit (f_k) une suite dans C_c telle que $0 \le f_k \le 1$ et $f_k \uparrow 1$ pour $k \to \infty$.



On décompose
$$\int \varphi\, d\mu_n - \int \varphi\, d\mu = \underbrace{\int \varphi\, d\mu_n - \int \varphi f_k\, d\mu_n}_{(A)} + \underbrace{\int \varphi f_k\, d\mu_n - \int \varphi f_k\, d\mu}_{(B)} + \underbrace{\int \varphi f_k\, d\mu - \int \varphi\, d\mu}_{(C)}$$

$(*)$ $\forall k$, (B) $\underset{n\to\infty}{\xrightarrow{(ii)}}$ $\quad (\varphi f_k \in C_c)$

$(*)$ $|(A)| \le \|\varphi\|_{L^\infty} \cdot (1 - \int f_k\, d\mu_n) \underset{n\to\infty}{\xrightarrow{(ii)}} \|\varphi\|_{L^\infty}(1 - \int f_k\, d\mu)$

$(*)$ $|(C)| \le \|\varphi\|_{L^\infty} \cdot (1 - \int f_k\, d\mu)$

Conclusion: $\forall k$,
$$\limsup_{n\to\infty} \left| \int \varphi\, d\mu_n - \int \varphi\, d\mu \right| \le 2\|\varphi\|_{L^\infty} \left(1 - \int f_k\, d\mu\right)$$
En prenant $k \to \infty$ on obtient $(\cdots) \longrightarrow 0$ (conv. monotone)
(iii) $\Longrightarrow$ (ii):
Supposons (iii). Soit $\varphi \in C_c$. Soit $\varphi_k \in H$ telle que $\|\varphi - \varphi_k\|_{L^\infty} \to 0$
Alors $\forall k$:

$$\limsup_{n\to\infty} \left| \int \varphi\, d\mu_n - \int \varphi\, d\mu \right| \le \limsup_{n\to\infty} \Bigg(\overbrace{\left| \int \varphi\, d\mu_n - \int \varphi_k\, d\mu_n \right|}^{\le \|\varphi - \varphi_k\|_{L^\infty}}$$
$$+ \underbrace{\left| \int \varphi_k\, d\mu_n - \int \varphi_k\, d\mu \right|}_{\underset{n\to\infty}{\longrightarrow}} + \underbrace{\left| \int \varphi_k\, d\mu - \int \varphi\, d\mu \right|}_{\le \|\varphi_k - \varphi\|_{L^\infty}} \Bigg) \le 2\|\varphi - \varphi_k\|_{L^\infty} \underset{k\to\infty}{\longrightarrow} 0$$

□

Attention: $X_n \xrightarrow{d} X$ n'implique pas en général que $\mathbb{P}(X_n \in B) \to \mathbb{P}(X \in B)$.
Exemple: $\delta(i)$: $\delta_{1/n} \xrightarrow{e} \delta_0$ mais $\delta_{1/n}(\{0\}) = 0 \not\to 1 = \delta_0(\{0\})$.
(Cf. convergence en variation totale, exercice 7.3)

Proposition 13 (Portmanteau)

Soient (μ_n), μ des mesures de probabilité sur $\mathbb{R}^d$. Sont équivalents:

(i) $\mu_n \xrightarrow{e} \mu$

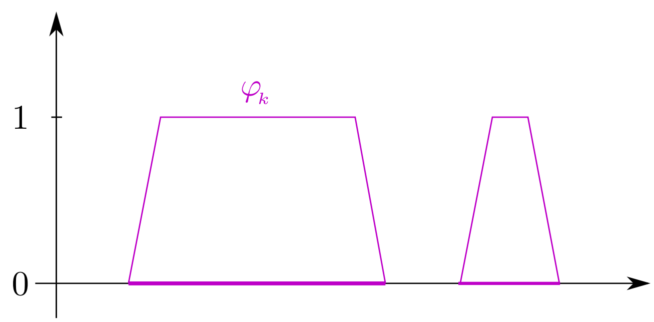
(ii) $\forall$ ouvert $G \subset \mathbb{R}^d$, $\liminf\limits_n \mu_n(G) \geq \mu(G)$

(iii) $\forall$ fermé $F \subset \mathbb{R}^d$, $\limsup\limits_n \mu_n(G) \leq \mu(F)$

(iv) $\forall B \in \mathcal{B}(\mathbb{R}^d)$ tel que $\mu(\partial B) = 0$: $\lim\limits_n \mu_n(B) = \mu(B)$

<u>*Démonstration:*</u>

<u>(i) $\Longrightarrow$ (ii):</u> Soit G ouvert. Alors il existe $\varphi_k \in C_b$, $k \in \mathbb{N}^*$, telles que $0 \leq \varphi_k \leq 1$ et $\varphi_k \nearrow 1_G$. Par exemple:



Donné par: $\varphi_k(x) := 1 \wedge \big(k \cdot \mathrm{dist}(x, G^c)\big)$

Alors comme $\varphi_k \leq 1_G$:

$$\begin{aligned}\liminf_n \mu_n(G) &\geq \sup_k \liminf_n \underbrace{\int \varphi_k \, d\mu_n}_{=\int \varphi_k \, d\mu} \\ &= \sup_k \int \varphi_k \, d\mu \underset{\text{Conv. Mono.}}{\leq} \mu(G)\end{aligned}$$

<u>(ii) $\Longleftrightarrow$ (iii):</u> Évident avec $F = G^c$.

<u>(ii) + (iii) $\Longrightarrow$ (iv):</u> Soit $B \in \mathcal{B}(\mathbb{R}^d)$.

$$\limsup_n \mu_n(B) \leq \limsup_n \mu_n(\overline{B}) \overset{(iii)}{\leq} \mu(\overline{B})$$

$$\liminf_n \mu_n(B) \geq \liminf_n \mu_n(B^\circ) \overset{(ii)}{\geq} \mu(B^\circ)$$

Rappel: $\partial B = \overline{B} \setminus B^\circ$
Si $\mu(\partial B)$ on a $\mu(\overline{B}) = \mu(B^\circ) = \mu(B)$ et on obtient (iv).

(iv) $\Longrightarrow$ (i): Soit $\varphi \subset C_b$ et supposons que $\varphi \geq 0$ (sinon on considère $\varphi = \varphi_+ - \varphi_-,\ \varphi_+, \varphi_- \geq 0$)
Avec $K := \sup_x \varphi(x)$ on a (Cf. Exercice 1.3):

$$\int \varphi(x)\ \mu(dx) = \int \int_0^K 1_{t \leq \varphi(x)}\ dt\ \mu(dx) \overset{\text{Fubini}}{=} \int_0^K \mu\left(E_t^\varphi\right)\ dt$$

Où: $E_t^\varphi := \{\ x \in \mathbb{R}^d \mid \varphi(x) \geq t\}$. De même:

$$\int \varphi(x)\ \mu_n(dx) = \int_0^K \mu_n\left(E_t^\varphi\right)\ dt$$

On a $\partial E_t^\varphi \subset \{x \in \mathbb{R}^d \mid \varphi(x) = t\}$

$\ulcorner$

φ continue $\Longrightarrow$ E_t^φ fermé et $\overset{\circ}{E_t^\varphi} \supset \{x \in \mathbb{R}^d \mid \varphi(x) > t\}$, qui est ouvert, $\subset E_t^\varphi$.
$\Longrightarrow$ $\partial E_t^\varphi = \overline{E_t^\varphi} \setminus \overset{\circ}{E_t^\varphi} \subset \{x \in \mathbb{R}^d \mid \varphi(x) = t\}$

$\lrcorner$

De plus, $\left\{t \in [0, K] \mid \mu\big(\{x \mid \varphi(x) = t\}\big) > 0\right\}$ est au plus dénombrable.

$\ulcorner$

$$\{\dots\} = \bigcup_{k \geq 1} \underbrace{\left\{t \mid \mu\big(\{x \mid \varphi(x) = t\}\big) \geq \frac{1}{k}\right\}}_{\#(\dots) \leq k \implies \text{ensemble fini}}$$

$\lrcorner$

Par (iv) $\mu_n\left(E_t^\varphi\right) \underset{n\to\infty}{\longrightarrow} \mu\left(E_t^\varphi\right)$ dt-p.p., et donc:

$$\int \varphi(x)\ \mu_n(dx) = \int_0^K \mu_n\left(E_t^\varphi\right)\ dt \underset{n\to\infty}{\overset{\text{conv. dom.}}{\longrightarrow}} \int_0^K \mu\left(E_t^\varphi\right)\ dt = \int \varphi(x)\ \mu(dx)$$

$\square$

Corollaire 13

Soient X_n, X des VA à valeurs dans $\mathbb{R}$. Alors

$$X_n \xrightarrow{d} X \iff F_{X_n}(x) \longrightarrow F_X(x) \quad \text{pour tout } x \text{ où } F \text{ est continue}$$

Démonstration:

$\Longrightarrow$: De Prop. 12 (iv)

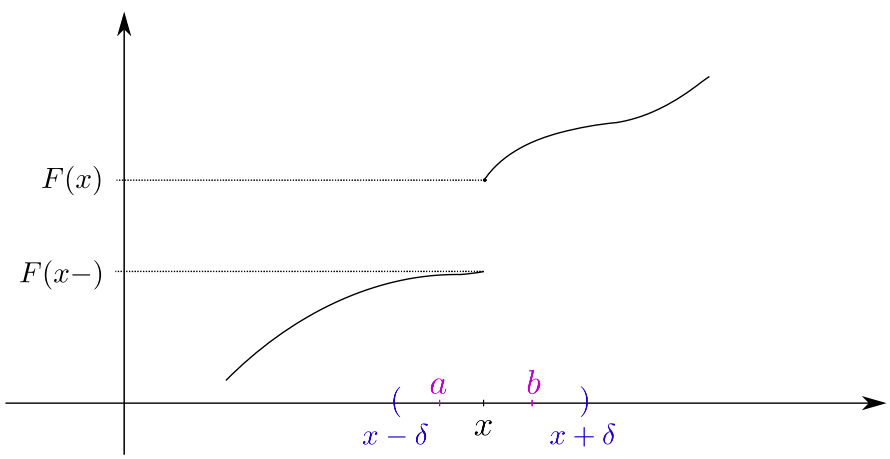
$\Longleftarrow$: Les points de discontinuité, D, de F forment un ensemble au plus dénombrable.

Pour $x \in D$ on a: $F(x-) := \lim_{y \nearrow x} F(y) < F(x)$. Donc il existe $q(x) \in \mathbb{Q} \cap \big(F(x-), F(x)\big)$. Par monotonie de F, $q: \ D \to \mathbb{Q}$ est injective.

Donc les points de continuité $\mathbb{R} \setminus D$ est dense dans $\mathbb{R}$.

Soit $x \in \mathbb{R}$ et $\varphi > 0$. Alors il existe $\delta > 0$ tel que:

$$F(x+\delta) \leq F(x) + \varepsilon \quad \text{et} \quad F(x-\delta) \geq F(x-) - \varepsilon$$



Soit $a \in [x-\delta, x]$ et $b \in [x, x+\delta]$ dans $\mathbb{R} \setminus D$. Alors on a

$$\lim_n F_{X_n}(a) = F(a) \geq F(x-\delta) \geq F(x-) \geq F(x-) - \varepsilon$$

$$\Longrightarrow \liminf_n F_{X_n}(x-) \geq \liminf_n F_{X_n}(a) \geq F(x-) - \varepsilon$$

Donc:

$$\liminf_n F_{X_n}(x-) \geq F(x-) \tag{$\star$}$$

De même:

$$\lim_n F_{X_n}(b) = F(b) \leq F(x) + \varepsilon \implies \limsup_n F_{X_n}(x) \leq \limsup_n F_{X_n}(b) \leq F(x) + \varepsilon$$

$$\Longrightarrow \limsup_n F_{X_n}(x) \leq F(x) \tag{$\star\star$}$$

Soient $\mu_n = \mathbb{P}_{X_n}$ et $\mu = \mathbb{P}_X$. De $(\star)$ et $(\star\star)$ on déduit que $\forall \alpha < \beta$ on a:

$$\liminf_n \mu_n\big((\alpha,\beta)\big) = \liminf_n \big(F_n(\beta-) - F_n(\alpha)\big) \geq F(\beta-) - F(\alpha) = \mu\big((a,b)\big)$$

Soit G un ouvert. Alors il existe des intervales ouverts disjoints $I_k,\ k \geq 1$, tels que $G = \bigcup_{k\geq 1} I_k$.

Pour $\varepsilon > 0$ fixé soit K tel que $\sum_{k>K} \mu(I_k) \leq \varepsilon$. Alors.

$$\begin{aligned}\liminf_n \mu_n(G) &\geq \liminf_n \mu_n\left(\bigcup_{k=1}^K I_k\right) = \liminf_n \sum_{k=1}^K \mu_n(I_k)\\ &\geq \sum_{k=1}^K \mu(I_k) \geq \mu(G) - \varepsilon\end{aligned}$$

Ce qui implique, par Prop. 12.(ii), que $\liminf_n \mu_n(G) \geq \mu(G)$, ce qui conclu.

$\square$

3.4 Fonctions caractéristiques

Définition 14

X une VA dans $\mathbb{R}^d$. La fonction caractéristique de X, ou la transformée de Fourier de $\mathbb{P}_X$, est:

$$\Phi_X(\xi) = \mathbb{E}\big[e^{i\xi\cdot X}\big] = \widehat{\mathbb{P}}_X(\xi) = \underbrace{\hat{\mu}(\xi) = \int e^{i\xi\cdot X} \;\; \mathbb{P}_X(dx)}_{\text{Notation d'analyste}} \quad \text{avec} \quad \xi \in \mathbb{R}^d$$

Note: $\Phi_X \in C_b(\mathbb{R}^d)$ par convergence dominée.

Proposition 15

Soit $X \in \mathbb{R}$ une VA gaussienne d'espérance 0 et de variance σ^2. Alors:

$$\Phi_X(\xi) = \exp\left(-\frac{\sigma^2}{2}\cdot\xi^2\right)$$

Démonstration:

$$\Phi_X(\xi) = \int \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right) \cdot \exp(i\xi x)\; dx$$

On peut supposer que $\sigma = 1$ et calculer:

$$f(\xi) := \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right) \cdot \exp(i\xi x)\; dx$$

On obtient:

$$\begin{aligned}
f'(\xi) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right) \cdot ix \cdot \exp(i\xi x)\; dx \\
&= \frac{1}{\sqrt{2\pi}} \cdot \int_{\mathbb{R}} (-i)\cdot \partial_x\Big[\exp\left(-\frac{x^2}{2}\right)\Big] \cdot \exp(i\xi x)\; dx \\
\overset{\text{IPP}}{\Longrightarrow} &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (-1)\exp\left(-\frac{x^2}{2}\right) \cdot \xi \cdot \exp(i\xi x)\; dx \\
&= -\xi \cdot f(\xi)
\end{aligned}$$

On obtient donc une equation différentielle:

$$\begin{cases} f(0) = 1 \\ f'(\xi) = -\xi \cdot f(\xi) \end{cases}$$

Et donc on obtient assez facilement que $f(\xi) = \exp\left(-\dfrac{\xi^2}{2}\right)$.

$\square$

Proposition 16

La transformée de Fourier $\hat{\mu}(\xi) = \displaystyle\int e^{i\xi x}\ \mu(dx)$ d'une mesure de probabilité μ sur $\mathbb{R}^d$ est caractérisée par μ, i.e. $\mu \to \hat{\mu}$ est injective.

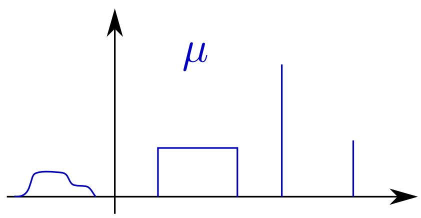
<u>*Démonstration:*</u>
Soit d'abord $d = 1$. Pour $\sigma > 0$ soit les fonctions:

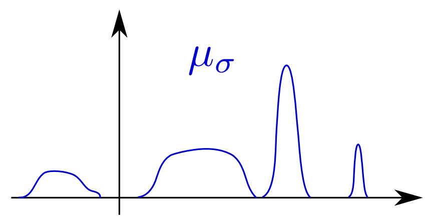
$$g_\sigma(x) := \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad \text{(Densité de } \mathcal{N}(0,\sigma^2))$$

$$f_\sigma(x) := g_\sigma * \mu(x) \overset{\text{def}}{=} \int g_\sigma(x-y)\ \mu(dy)$$

Et: $\mu_\sigma(dx) := f_\sigma(x)\ dx$

On peut voir la mesure μ_σ comme une sorte de lissage de la mesure μ donnée:





Il suffit de montrer que:

(i) μ_σ est détérminé par $\hat{\mu}$

(ii) $\mu_\sigma \xrightarrow[x^3]{\text{e}} \mu$

$\boxed{\text{(i)}}$:

$$\sigma\sqrt{2\pi} g_\sigma(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \overset{\text{Prop. 15}}{=} \int e^{i\xi x} \cdot g_{\frac{1}{\sigma}}(\xi)\ d\xi$$

donc:

$$\begin{aligned} f_\sigma(x) &= \int g_\sigma(x-y)\ \mu(dy) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int\int e^{i\xi(x-y)} g_{\frac{1}{\sigma}}(\xi)\ d\xi\ \mu(dy) \\ \overset{\text{Fubini}}{\Longrightarrow} &= \frac{1}{\sigma\sqrt{2\pi}} \int e^{i\xi x} g_{\frac{1}{\sigma}}(\xi) \int e^{-i\xi y}\ \mu(dy)\ d\xi \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int e^{i\xi x} g_{\frac{1}{\sigma}}(\xi)\ \hat{\mu}(-\xi)\ d\xi \end{aligned} \qquad (\star)$$

$\boxed{\text{(ii)}}$: Pour $\varphi \in C_b$ on a

$$\int \varphi(x)\ \mu_\sigma(dx) = \int \varphi(x) \int g_\sigma(x-y)\ \mu(dy)\ dx \overset{\text{Fubini}}{=} \int g_\sigma * \varphi(y)\ \mu(dy)$$

Par convergence dominée, on a $g_\sigma * \varphi(y) \underset{\sigma\to 0}{\longrightarrow} \varphi(y)$:

$$g_\sigma * \varphi(y) = \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \cdot \varphi(y-x)\, dx = \int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \cdot \varphi(y-\sigma x)\, dx \longrightarrow \varphi(y)$$

Donc

$$\int \varphi(x)\; \mu_\sigma(dx) \longrightarrow \int \varphi(y)\; \mu(dy)$$

Pour $d > 1$: Même argument avec:

$$g_\sigma^{(d)}(x_1, \ldots, x_d) := \prod_{i=1}^{d} g_\sigma(x_i)$$

$\square$

Petite remarque: Formellement, avec $\sigma = 0$, on a pour $\mu(dx) = f(x)\, dx$ et $\hat{f} = \hat{\mu}$,

$$f(x) = \int e^{i-x\xi}\; \hat{f}(\xi)\; d\xi$$

Comparer avec:

$$\hat{f}(x) \stackrel{\text{def}}{=} \int e^{ix\xi} \cdot f(x)\; dx$$

Proposition 17

Soient μ_n, μ des mesures de probabilités sur $\mathbb{R}^d$. Alors:

$$\mu_n \xrightarrow[n\to\infty]{\text{e}} \mu \iff \hat{\mu}_n(\xi) \longrightarrow \hat{\mu}(\xi) \qquad \forall \xi \in \mathbb{R}^d$$

Démonstration:

$\Longrightarrow$: Évident: La transformée de Fourier d'une mesure de probabilité est l'intégrale de cette mesure par rapport à la fonction e^{xi} qui est une fonction continue et bornée. Donc cela suit directement de la définition de convergence étroite.

$\Longleftarrow$: On traite uniquement le cas $d = 1$ ($d > 1$ est très similaire).
Soit $\hat{\mu}_n(\xi) \longrightarrow \hat{\mu}(\xi) \quad \forall \xi \in \mathbb{R}$. Soit $\varphi \in C_c$. Alors:

$$\int g_\sigma * \varphi\; d\mu \stackrel{\text{Fubini}}{=} \int \varphi(x) \cdot (g_\sigma * \mu)(x)\; dx \stackrel{(\star)}{=} \int \varphi \frac{1}{\sigma\sqrt{2\pi}} \int e^{i\xi x} g_{\frac{1}{\sigma}}(\xi) \hat{\mu}(-\xi)\; d\xi\; dx$$

et pareil pour μ_n. Par convergence dominée:

$$\int e^{i\xi x} g_{\frac{1}{\sigma}}(\xi) \hat{\mu}_n(-\xi)\; d\xi \quad \xrightarrow[n\to\infty]{} \quad \int e^{i\xi x} g_{\frac{1}{\sigma}}(\xi) \hat{\mu}(-\xi)\; d\xi \qquad \forall x$$

Et donc, à nouveau par convergence dominée:

$$\int g_\sigma * \varphi\; d\mu_n \quad \xrightarrow[n\to\infty]{} \quad \int g_\sigma * \varphi\; d\mu \qquad \forall \varphi \in C_c \tag{$\star\star$}$$

Finalement, soit $H := \{g_\sigma * \varphi \mid \sigma > 0 \quad \text{et} \quad \varphi \in C_c\}$. Alors l'adhérence de G par rapport à $\|\cdot\|_{L^\infty}$ contient C_c.

$$\psi \in C_c \implies \|g_\sigma * \psi - \psi\|_{L^\infty} = \sup_x \left| \int \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right) \cdot \big(\psi(x-y) - \psi(x)\big)\, dy \right|$$
$$= \sup_x \left| \int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \big(\psi(x-\sigma y) - \psi(x)\big)\, dy \right|$$

Soit $\varepsilon > 0$ et $K > 0$ tel que:

$$\int_{|y|\geq K} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)\, dy \leq \frac{\varepsilon}{\|\psi\|_{L^\infty}}$$

Alors par continuité uniforme de ψ:

$$\|g_\sigma * \psi - \psi\|_{L^\infty} \leq \sup_x \left| \int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \big(\psi(x-\sigma y) - \underbrace{\psi(x)}_{|\cdot|\leq K\cdot\sigma}\big)\, dy \right| + 2\varepsilon \quad \underset{\sigma\to 0}{\longrightarrow} \quad 2\varepsilon$$

Par $(\star\star)$, on a $\int \psi\, d\mu_n \underset{n\to\infty}{\longrightarrow} \int \psi\, d\mu \quad \forall \psi \in H$. On conclut par la Proposition 12.

□

3.5 Le théorème central limite

Soient $(X_n)_{n\geq 1}$ une suite de VA réelles indépendantes de même loi dans L^1. La loi des grands nombres dit que:

$$\frac{1}{n}(X_1 + \cdots + X_n) \quad \xrightarrow[n\to\infty]{\text{p.s.}} \quad \mathbb{E}\big[X_1\big]$$

Question: Quelle est la vitesse de convergence ? i.e. l'ordre de grandeur (en termes de n) de l'expression $\frac{1}{n}(X_1 + \cdots + X_n) - \mathbb{E}[X_1]$. Pour $X_n \in L^2$, la réponse est facile à deviner:

$$\mathbb{E}\Big[\big(X_1 + \cdots + X_n - n\cdot\mathbb{E}[X_1]\big)^2\Big] = \mathrm{Var}(X_1 + \cdots + X_n) = n\cdot\mathrm{Var}(X_1)$$

Donc $\dfrac{X_1 + \cdots + X_n - n\cdot\mathbb{E}[X_1]}{\sqrt{n}}$ est d'ordre 1.

Résultat plus précis qui identifie même la loi des fluctuations:

Proposition 18: Théorème central limite

Soit $(X_n)_{n\geq 1}$ VA réelles indépendantes de même loi dans L^2 avec variance σ^2. Alors:

$$\frac{1}{\sqrt{n}}\Big((X_1 + \cdots + X_n) - n\cdot\mathbb{E}[X_1]\Big) \quad \xrightarrow[n\to\infty]{d} \quad \mathcal{N}(0,\sigma^2)$$

<u>Démonstration:</u>
On peut supposer que $\mathbb{E}[X_1] = 0$ (sinon on remplace X_n par $X_n - \mathbb{E}[X_n]$).
On utilise que pour $X \in L^2$ on a:

$$\Phi_X(\xi) = 1 + i\cdot\xi\cdot\mathbb{E}[X] - \frac{1}{2}\xi^2\cdot\mathbb{E}[X^2] + o(\xi^2) \tag{$\star$}$$

$$\Phi_X'(\xi) = i\cdot\mathbb{E}\big[X\exp(i\xi X)\big] \quad \text{et} \quad \Phi_X''(\xi) = -\mathbb{E}\big[X^2\exp(i\xi X)\big]$$

Donc $X\exp(i\xi X)$ et $X^2\exp(i\xi X)$ sont L^1, et comme les Φ_X' et Φ_X'' sont continue (On peut facilement vérifier cela). Donc $\Phi_X \in C^2(\mathbb{R})$ par convergence dominée. Utiliser Taylor ensuite.

Avec $Z_n = \dfrac{1}{\sqrt{n}}\big(X_1 + \cdots + X_n\big)$ on a:

$$\begin{aligned}
\Phi_{Z_n}(\xi) &= \mathbb{E}\Big[\exp\left(i\xi\cdot\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\Big] = \mathbb{E}\Big[\exp\left(i\xi\cdot\frac{X_1}{\sqrt{n}}\right)\Big]^n \\
&= \Phi_{X_1}\left(\frac{\xi}{\sqrt{n}}\right)^n \overset{(\star)}{=} \left(1 - \frac{\xi^2}{2n}\sigma^2 + o\big(\frac{\xi^2}{n}\big)\right)^n \xrightarrow[n\to\infty]{} \exp\left(-\frac{\sigma^2\xi^2}{2}\right) \quad \forall\xi\in\mathbb{R} \\
&\overset{\text{Par Prop 15}}{\Longrightarrow} \exp\left(-\frac{\sigma^2\xi^2}{2}\right) = \Phi_{\mathcal{N}(0,\sigma^2)}(\xi)
\end{aligned}$$

Et on conclu donc par Proposition 17.

□

Remarque: Notons que cette proposition ne répond pas a la question de vitesse de convergence.