Igor Kortchenski

Cours 3 : Propriétés de v.a. $\perp\!\!\!\perp$

Plan :

1) Construction d'une suite de v.a. $\perp\!\!\!\perp$
2) Loi du 0-1 de Kolmogorov
3) Sommes de v.a $\perp\!\!\!\perp$
4) Semi-groupe de convolution
5) Loi de grands nombres

1) Construction d'une suite de v.a. ⫫

Nous allons voir qu'on peut utiliser le développement diadique (vu au cours 2) pour construire une suite infinie de v.a. réelles i.i.d.

i.i.d = indépendantes et identiquement distribuées (de même loi)

(on peut aussi le faire dans un cadre plus général que v.a. réelles mais cela dépasse le cadre de ce cours)

Rappel du cours 2

$(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathcal{B}([0,1[), \lambda)$

avec λ : mesure de Lebesgue

$\omega \in [0,1[$,

$$\omega = \sum_{n=1}^{\infty} \frac{X_n(\omega)}{2^n}$$

avec $X_n(\omega) = \lfloor 2^n \omega \rfloor - 2 \cdot \lfloor 2^{n-1} \omega \rfloor$

On a vu que la suite $(X_n)_{n \geq 1}$ est une suite de v.a à valeurs dans $\{0,1\}$, i.i.d de loi :

$\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$,

c'est-à-dire Bernoulli$(\frac{1}{2})$

Idée : utiliser cette suite pour construire plein d'autres v.a.

Étape 1 : Soit $\varphi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ une injection, par exemple
$$\varphi(i,j) = 2^i(2j+1)$$
On pose
$$Y_{i,j} = X_{\varphi(i,j)}$$
pour $i,j \geq 0$.
Clairement les $(Y_{i,j})_{i,j \geq 0}$ sont iid Bernoulli(1/2)

Étape 2 On pose pour $i \geq 0$,
$$U_i = \sum_{j=1}^{\infty} \frac{Y_{i,j}}{2^j}$$

$$
\begin{array}{llll}
Y_{1,1} & Y_{1,2} & Y_{1,3} & \cdots \rightsquigarrow U_1 \\
Y_{2,1} & Y_{2,2} & Y_{2,3} & \cdots \rightsquigarrow U_2 \\
Y_{3,1} & Y_{3,2} & Y_{3,3} & \cdots \rightsquigarrow U_3
\end{array}
$$

Remarque les v.a. $(U_i)_{i\geq 1}$ sont iid de loi uniforme sur $[0,1]$

$$\boxed{U_i = \sum_{j=1}^{\infty} \frac{Y_{i,j}}{2^j}}$$

Preuve • U_i est $\sigma(Y_{i,j} : j \geq 1)$ mesurable, donc les $(U_i)_{i\geq 1}$ sont $\perp\!\!\!\perp$ par une extension du principe des coalitions à des familles infinies (cf cours 2)

Ensuite, pour $p \geq 1$ on définit

$$W_i^{(p)} = \sum_{j=1}^{p} \frac{Y_{i,j}}{2^j}, \text{ qui}$$

a la même loi que

$$W^{(p)} = \sum_{j=1}^{p} \frac{X_j}{2^j} \text{ car}$$

$$(Y_{i,1}, \dots, Y_{i,p}) \overset{loi}{=} (X_1, \dots, X_p)$$

On va passer à la limite quand $p \to \infty$. Pour cela on prend $f : \mathbb{R} \to \mathbb{R}$ continue bornée. On a

$$\mathbb{E}[f(W_i^{(p)})] = \mathbb{E}[f(W^{(p)})].$$

On utilise le théorème de convergence dominée :

$$\mathbb{E}[f(U_i^{(p)})] \xrightarrow[p\to\infty]{} \mathbb{E}[f(U_i)]$$

$$\mathbb{E}[f(U^{(p)})] \xrightarrow[p\to\infty]{} \mathbb{E}[f(U)]$$

avec $U(\omega) = \omega$ uniforme sur $[0,1]$.

Donc $\forall f : \mathbb{R} \to \mathbb{R}_+$ continue bornée,

$$\mathbb{E}[f(U_i)] = \mathbb{E}[f(U)]$$

Donc $U_i \overset{\text{loi}}{=} U$.

D'où U_i est uniforme

$\sim$

"$=$" $\neq$ "$\overset{\text{loi}}{=}$"

- Dire $X = Y$ c'est dire $\forall \omega \in \Omega, \ X(\omega) = Y(\omega)$
- Dire $X \overset{\text{loi}}{=} Y$ c'est dire $\mathbb{P}_X = \mathbb{P}_Y$, c'est à dire $\forall A \ \ \mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$

<u>Exemple</u> Si $X = \text{Bernoulli}(\frac{1}{2})$ et $Y = 1 - X$. Alors

$X \overset{\text{loi}}{=} Y$ mais $\forall \omega \in \Omega, X(\omega) \neq Y(\omega)$

Lemme Soit μ une mesure de probabilité sur $\mathbb{R}$. Soit

$$F_\mu(x) = \mu(]-\infty, x])$$

sa fonction de répartition.

On pose

$$F_\mu^{-1}(y) = \inf\{x \in \mathbb{R} : F_\mu(x) \geq y\}$$

On pose $Z_i = F_\mu^{-1}(U_i)$

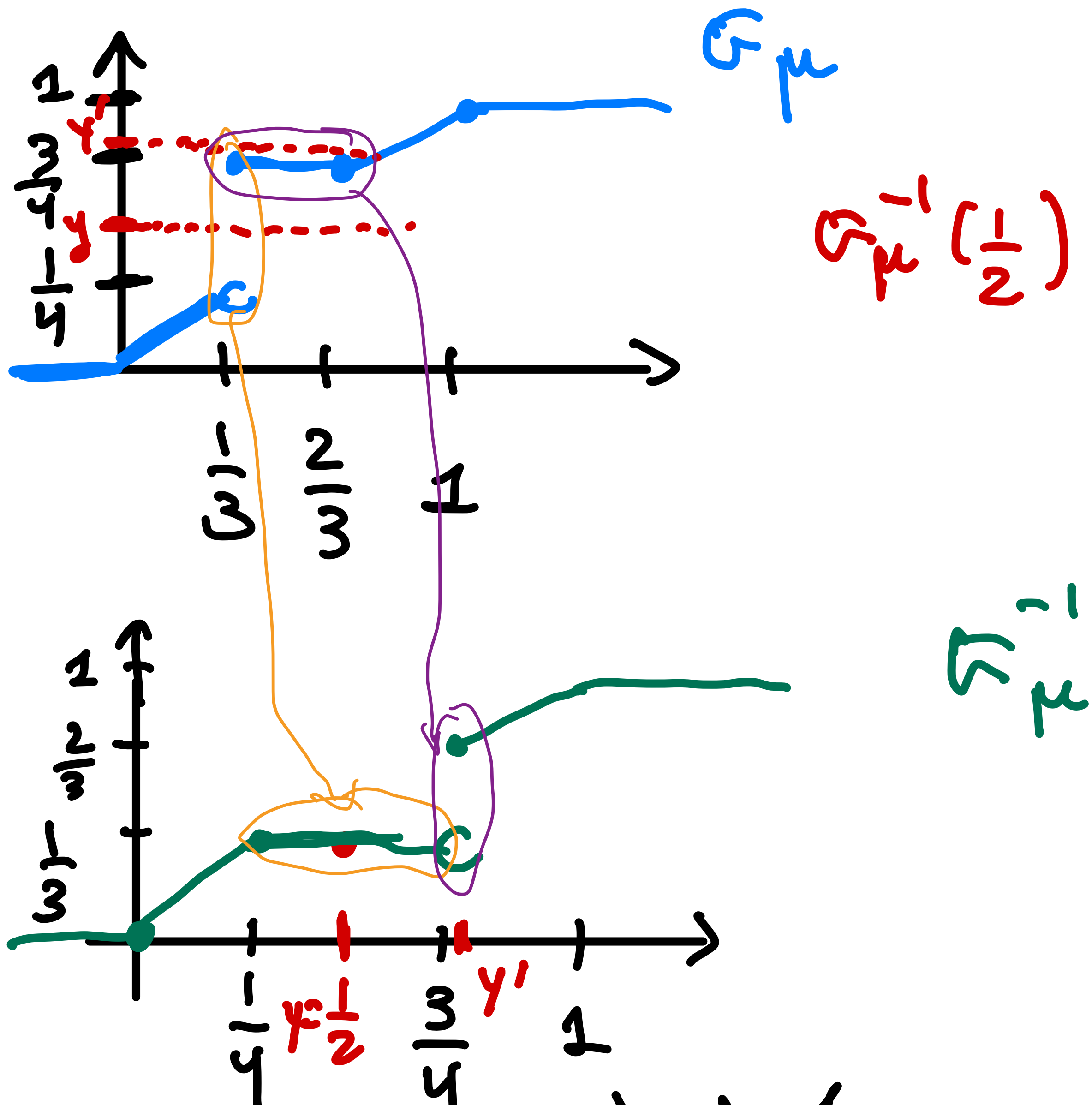
Alors $(Z_i)_{i \geq 1}$ sont iid de loi μ.

Avant la preuve, quelques exemples.

⚠ F_μ n'est pas toujours

bijectif : σ_μ^{-1} est appelé <u>inverse généralisé</u>

•



$\sigma_\mu^{-1}\left(\frac{1}{2}\right) = \frac{1}{3}$

• Si $\sigma_\mu(x) = (1-e^{-\lambda x})\mathbb{1}_{x \geq 0}$, fonction de répartition d'une $\mathcal{E}xp(\lambda)$

$$1-e^{-\lambda x} = y \iff x = -\frac{1}{\lambda}\ln(1-y)$$

Ainsi, si U est une v.a. uniforme sur $[0,1]$, $-\frac{1}{\lambda}\ln(1-U)$ suit une loi $\mathcal{E}xp(\lambda)$.

Lemme Soit μ une mesure de probabilité sur $\mathbb{R}$. Soit

$$F_\mu(x) = \mu(]-\infty, x])$$

sa fonction de répartition.

On pose

$$F_\mu^{-1}(y) = \inf\{x \in \mathbb{R} : F_\mu(x) \geq y\}$$

On pose $Z_i = F_\mu^{-1}(U_i)$

Alors $(Z_i)_{i\geq 1}$ sont iid de loi μ.

Preuve du lemme

D'après le principe de composition, les v.a. $(Z_i)_{i\geq 1}$ sont iid.
Il suffit donc de montrer que

$$Z_1 \overset{\text{loi}}{=} \mu.$$

Pour cela, on remarque que pour $x \in \mathbb{R}$ et $y \in]0,1[$,

$$\boxed{F_\mu^{-1}(y) \leq x \underset{(*)}{\Longleftrightarrow} y \leq F_\mu(x)}$$

Ainsi $\forall a \in \mathbb{R}$,

$$\mathbb{P}(Z_1 \leq a) = \mathbb{P}(F_\mu^{-1}(U_1) \leq a)$$

$$= \mathbb{P}(U_1 \leq F_\mu(a)) \text{ par } (*)$$

$= F_\mu(a)$ car U_1 est uniforme sur

$[0,1]$.

Ainsi Z_1 et μ ont même fonction de répartition, donc même loi.

Résumé

U uniforme sur $[0,1]$

$\Downarrow$

1 suite iid Bernoulli$(1/2)$

$\Longrightarrow$ suite de v.a. iid uniformes $[0,1]$

$\Downarrow$

suite de v.a iid de loi μ

2) Loi du 0-1 de Kolmogorov

Soit $(X_n)_{n\geq 1}$ des v.a. On note

$\mathcal{B}_n = \sigma(X_k : k \geq n)$

et $\mathcal{B}_\infty = \bigcap_{n\geq 0} \mathcal{B}_n$, appelée

tribu queue.

Intuition $\mathcal{B}_\infty$ contient toute l'information des $(X_n)_{n\geq 1}$ qui ne dépend pas d'un nombre fini de v.a.

Exemple : Si $(X_n)_{n\geq 1}$ sont des v.a. réelles,

$\{\sup_n S_n = +\infty\} \in \mathcal{B}_\infty$ où

$S_n = X_1 + \dots + X_n$.

En effet, $\forall p \geq 1$

$\boxed{\mathcal{B}_n = \sigma(X_k : k \geq n) \text{ et } \mathcal{B}_\infty = \bigcap_{n \geq 0} \mathcal{B}_n,}$

$$\{\sup_n S_n = +\infty\} = \{\sup_n (X_p + X_{p+1} + \dots + X_{p+n}) = \infty\} \in \mathcal{B}_p$$

Théorème (loi du 0-1 de Kolmogorov)

On suppose que les v.a. $(X_n)_{n \geq 1}$ sont $\perp\!\!\!\perp$. Alors $\forall B \in \mathcal{B}_\infty$, $\mathbb{P}(B) = 0$ ou 1.

On dit que la tribu $\mathcal{B}_\infty$ est triviale

<u>Preuve</u> : Soit $\mathcal{D}_n = \sigma(X_1, \dots, X_{n-1})$

Alors $\mathcal{D}_n \perp\!\!\!\perp \mathcal{B}_n$ (cf cours 2)

Donc $\mathcal{D}_n \perp\!\!\!\perp \mathcal{B}_\infty$ car $\mathcal{B}_\infty \subset \mathcal{B}_n$

Ainsi, $\forall A \in \bigcup_{n=1}^{\infty} \mathcal{D}_n$, $\forall B \in \mathcal{B}_\infty$,

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

Or $\bigcup_{n=1}^{\infty} \mathcal{D}_n$ est un π-système

générateur de $\sigma(X_i : i \geq 1)$

Donc

$$\sigma(X_i : i \geq 1) \perp\!\!\!\perp \mathcal{B}_\infty$$

or $\mathcal{B}_\infty \subset \sigma(X_i : i \geq 1)$

Donc $\mathcal{B}_\infty \perp\!\!\!\perp \mathcal{B}_\infty$!

Donc $\forall A \in \mathcal{B}_\infty$,

$$\begin{array}{ccc} \mathbb{P}(A \cap A) & = & \mathbb{P}(A)\,\mathbb{P}(A) \\ \| & & \| \\ \mathbb{P}(A) & & \mathbb{P}(A)^2 \end{array}$$

Donc $\mathbb{P}(A) = 0$ ou 1.

<u>Remarque</u> Une v.a. réelle mesurable par rapport à une tribu triviale est constante p.s.
(exercice ; indication : sa fonction de répartition vaut 0 ou 1)

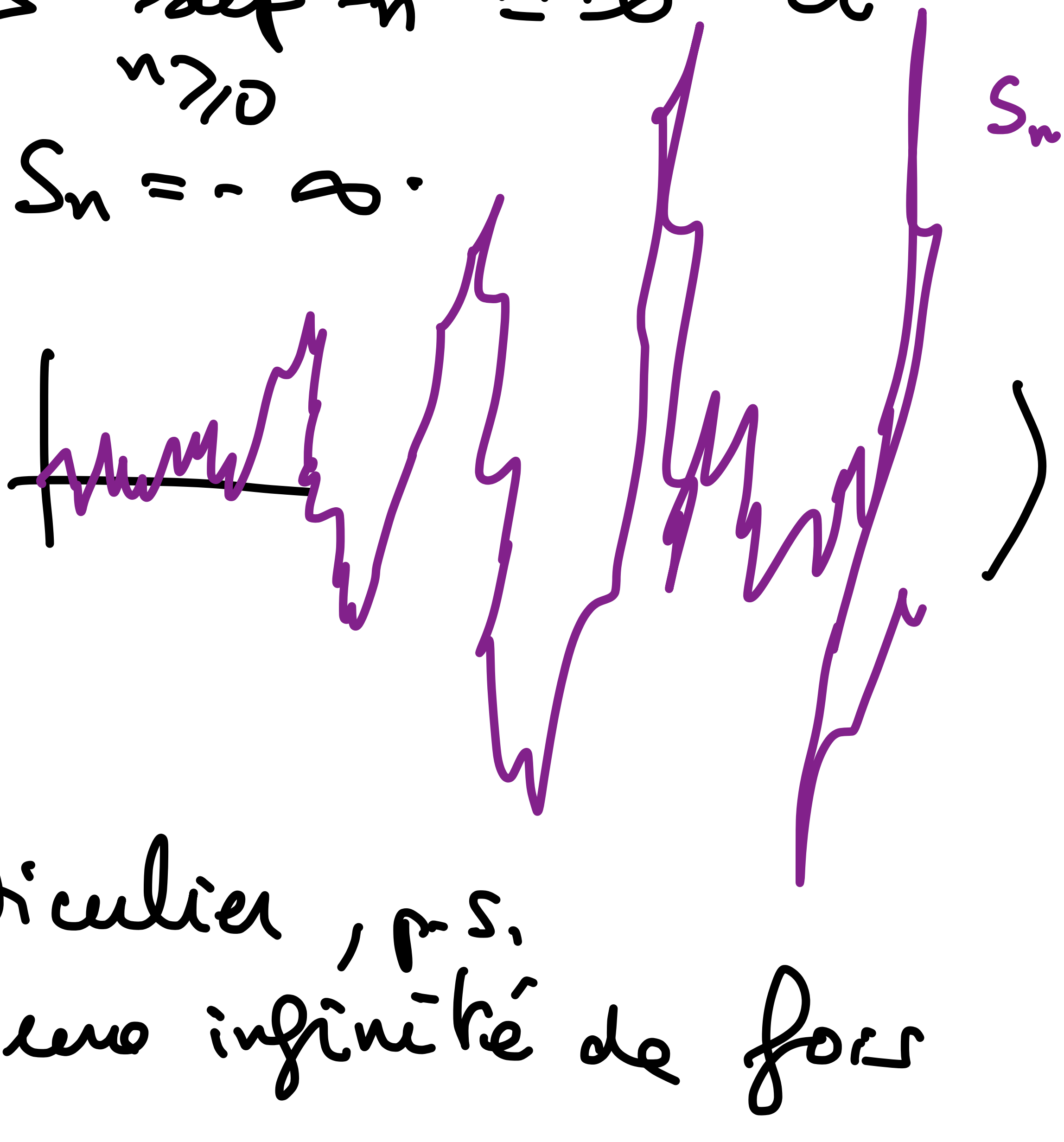
Application : Soit $(X_n)_{n\geq 1}$ iid

avec $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$

On pose $S_n = X_1 + \cdots + X_n$

Alors p.s $\sup_{n\geq 0} S_n = \infty$ et

p.s. $\inf_{n\geq 0} S_n = -\infty$.

(image 

)

En particulier, p.s.

$S_n = 0$ une infinité de fois

<u>Preuve</u> On montre que $\forall p \geq 1$

$$\mathbb{P}\left(-p \leq \inf_n S_n \leq \sup_n S_n \leq p\right) = 0 \qquad (*)$$

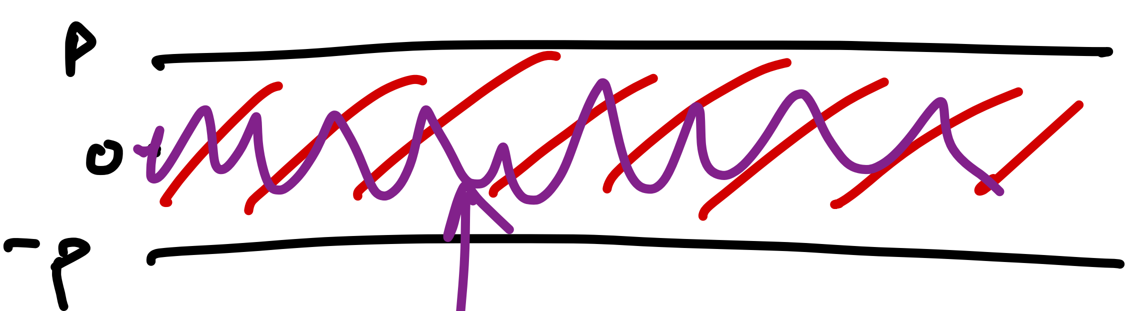
Pour cela, soit $k > 2p$ fixé

on remarque que

(k fois +1 à la suite)

$$\bigcup_{j=0}^{\infty} \{X_{jk+1} = +1, X_{jk+2} = +1, \dots, X_{jk+k} = 1\}$$

$$\subset \left\{-p \leq \inf_n S_n \leq \sup_n S_n \leq p\right\}^c$$



$\{-p \leq \inf_n S_n \leq \sup_n S_n \leq p\}$

Par BC2 (cf fin du cours 2)
on a

$$\mathbb{P}\left(\bigcup_{j=0}^{\infty} \{ X_{jk+1} = +1, \ldots, X_{jk+k} = +1 \} \right) = 1.$$

D'où (*)

En faisant $p \to \infty$, on obtient

$$\mathbb{P}\left(\{ \inf S_n > -\infty \} \cup \{ \sup S_n < \infty \} \right) = 0$$

Donc $\mathbb{P}(\inf S_n = -\infty)$
$+ \mathbb{P}(\sup S_n = +\infty) \geq 1$ (**)

Or par symétrie

$$\mathbb{P}(\inf S_n = -\infty) = \mathbb{P}(\sup S_n = \infty)$$

Or d'après la loi de 0-1 de Kolmogorov, comme $\{\inf S_n = -\infty\}$, $\{\sup S_n = +\infty\} \in \mathcal{B}_\infty$, leur probabilité vaut 0 ou 1.

Donc par (**) cette probabilité vaut 1 !

$\sim$

3) Sommes de v.a. $\perp\!\!\!\perp$

Si μ, ν sont deux mesures de probabilité sur $\mathbb{R}^d$, on note $\mu * \nu$ la mesure image de $\mu \otimes \nu$ par $(x,y) \longmapsto x+y$.

c'est-à-dire $\forall f : \mathbb{R}^d \to \mathbb{R}_+$ mesurable,

$$\int_{\mathbb{R}^d} f(z)\, \mu * \nu(dz) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x+y)\, \mu(dx)\, \nu(dy)$$

Proposition Soient X, Y deux v.a $\perp\!\!\!\perp$ dans $\mathbb{R}^d$

(1) La loi de $X+Y$ est $\mathbb{P}_X * \mathbb{P}_Y$

En particulier, si X a pour

densité p_X et Y a pour densité p_Y, alors $X+Y$ a pour densité

$$p_X * p_Y(z) = \int_{\mathbb{R}^d} p_X(x)\, p_Y(z-x)\,dx$$

produit de convolution

② Si X et Y sont dans L^2, alors $K_{X+Y} = K_X + K_Y$ où

$$(K_Z)_{i,j} = \mathrm{Cov}(Z_i, Z_j) = \mathbb{E}[Z_i Z_j] - \mathbb{E}[Z_i]\mathbb{E}[Z_j]$$

avec $Z = (Z_1, \dots, Z_d)$

<u>Preuve</u> : ① Comme $X \perp\!\!\!\perp Y$ on sait que $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$. On utilise le principe de la fonction muette :

Soit $f : \mathbb{R}^d \to \mathbb{R}_+$ mesurable

$\mathbb{E}[f(X+Y)]$

$= \int f(x+y)\, \mathbb{P}_{(X,Y)}(dx\,dy)$ (transfert)

$= \int f(x+y)\, \mathbb{P}_X(dx)\, \mathbb{P}_Y(dy)$

$= \int f(z)\, \mathbb{P}_X * \mathbb{P}_Y(dz)$

par définition de $*$.

Si X et Y ont une densité on a

$\mathbb{E}[f(X+Y)]$

$= \int f(x+y)\, p_X(x)\, p_Y(y)\, dx\, dy$

$= \int f(z) \left(\int p_X(x)\, p_Y(z-x)\, dx \right) dz$

(changement de variable
$x+y = z$
$x = x$)

donc $\mathbb{P}_X * \mathbb{P}_Y(dz)$

$$= \underbrace{\left(\int p_X(x)\, p_Y(z-x)\,dx\right)}_{\text{densité de } \mathbb{P}_X * \mathbb{P}_Y.} dz$$

② Si

$X = (X_1, \dots, X_d)$

$Y = (Y_1, \dots, Y_d)$

Par $\perp\!\!\!\perp$, on a $\mathrm{Cov}(X_i, Y_j) = 0$ $\forall\ 1 \leq i,j \leq d$. Par linéarité,

$\mathrm{Cov}(X_i + Y_i, X_j + Y_j)$

$= \mathrm{Cov}(X_i, X_j) + \mathrm{Cov}(Y_i, Y_j)$

Donc $K_{X+Y} = K_X + K_Y$.

∞

4) Semigroupes de convolution

Soit $I = \mathbb{N}$ ou $I = \mathbb{R}_+$.

Définition Soit $(\mu_t)_{t \in I}$ une famille de mesures de probabilité sur $\mathbb{R}$ ou $\mathbb{R}^d$. On dit que $(\mu_t)_{t \in I}$ est un semi-groupe de convolution si :

$$\begin{cases} \mu_0 = \delta_0 \\ \mu_t * \mu_{t'} = \mu_{t+t'} \quad \forall t, t' \in I \end{cases}$$

Interprétation probabiliste

- $X_0 = 0$ p.s
- si $\text{loi}(X_t) = \mu_t$, $\text{loi}(X_{t'}) = \mu_{t'}$

si $X_t \perp\!\!\!\perp X_{t'}$, alors

$$\text{loi}(X_{t+t'}) = \mu_{t+t'}$$

On verra plus tard en lien avec les fonctions caractéristiques.

<u>Exemple</u> $I = \mathbb{N}$,

$\mu_n =$ loi Binomiale (n, p) avec p fixé.

(on utilise l'interprétation probabiliste, somme de n v.a. i.i.d Bernoulli(p)).

5) Lois des grands nombres

Théorème (loi faible des grands nombres, version L^2)

Soit $(X_n)_{n\geq 1}$ des v.a. réelles i.i.d. Si $\mathbb{E}[X_1^2] < \infty$, on a

$$\frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{L^2} \mathbb{E}[X_1]$$

Preuve. Par linéarité,

$$\mathbb{E}\left[\frac{1}{n}(X_1 + \dots + X_n)\right] = \mathbb{E}[X_1].$$

Donc

$$\mathbb{E}\left[\left(\frac{1}{n}(X_1+\dots+X_n)-\mathbb{E}[X_1]\right)^2\right]$$

$$=\mathrm{Var}\left(\frac{X_1+\dots+X_n}{n}\right)$$

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

$$=\frac{1}{n^2}\left(\mathrm{Var}(X_1)+\dots+\mathrm{Var}(X_n)\right) \text{ par } \perp\!\!\!\perp$$

$$=\frac{1}{n}\mathrm{Var}(X_1)\xrightarrow[n\to\infty]{}0$$

Remarque ceci implique

$\forall \varepsilon>0,$

$$\mathbb{P}\left(\left|\frac{1}{n}(X_1+\dots+X_n)-\mathbb{E}[X_1]\right|>\varepsilon\right)\xrightarrow[n\to\infty]{}0$$

(cv L^2 $\Rightarrow$ cv en probabilité : vu plus tard)

Remarque la preuve montre que le résultat reste vrai avec des hypothèses bien plus faibles :

- on peut remplacer "les X_i ont même loi" par " $\mathbb{E}[X_1] = \mathbb{E}[X_n]$ $\forall n$" et $(\mathbb{E}[X_i^2])_{i \geq 1}$ borné
- on peut remplacer "$(X_i)_{i\geq 1}$ sont $\perp\!\!\!\perp$" par "$\mathrm{Cov}(X_i, X_j) = 0$ pour $i \neq j$"

D'un point de vue probabiliste, il est plus significatif d'avoir une convergence p.s, c'est-à-dire en dehors d'un ensemble de probabilité nulle ; on parle de loi forte

Donnons un premier résultat en ce sens. On verra la prochaine fois un résultat plus général mais avec une preuve bien plus délicate.

Théorème (loi forte des grands nombres, version L^4)

Soit $(X_i)_{i\geq 1}$ v.a. réelles iid avec $\mathbb{E}[X_1^4]<\infty$. Alors p.s. $\frac{1}{n}(X_1+\dots+X_n) \xrightarrow[n\to\infty]{} \mathbb{E}[X_1]$

<u>Preuve</u> : Quitte à remplacer X_i par $X_i - \mathbb{E}[X_i]$ on peut supposer $\mathbb{E}[X_1]=0$ (<u>centrées</u>)

<u>Idée</u> Montrer que

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

En effet, si $(*)$ est vrai, on a alors par Fubini

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

Donc p.s. $\displaystyle\sum_{n=1}^{\infty}\left(\frac{X_1+\cdots+X_n}{n}\right)^4<\infty$

Donc p.s $\displaystyle\left(\frac{X_1+\cdots+X_n}{n}\right)^4 \xrightarrow[n\to\infty]{} 0$

Donc p.s. $\displaystyle\frac{X_1+\cdots+X_n}{n} \xrightarrow[n\to\infty]{} 0$

Pour montrer $(*)$ on écrit

$$\mathbb{E}\left[\left(\frac{X_1+\cdots+X_n}{n}\right)^4\right]=$$

$$\frac{1}{n^4} \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}]$$

On utilise l'$\perp\!\!\!\perp$ et le fait que $\mathbb{E}[X_1] = 0$: on voit que les seuls termes de la somme qui sont $\neq 0$ sont ceux où toute valeur parmi (i_1, i_2, i_3, i_4) est prise 2 ou 4 fois.

(par exemple $\mathbb{E}[X_2 X_1 X_2 X_2]$
$= \mathbb{E}[X_2^3]\, \mathbb{E}[X_1] = 0$)

Ainsi

$$\sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}]$$

$$= n\,\mathbb{E}[X_1^4] + 3n(n-1)\,\mathbb{E}[X_1^2 X_2^2]$$

→ terme combinatoire

choisir (i_1, i_2, i_3, i_4) de sorte qu'il y ait 2 valeurs a, b prises 2 fois c'est :

- choisir $\{a, b\}$: $\binom{n}{2} = \frac{n(n-1)}{2}$ choix
- choisir l'ordre :

(a, a, b, b) (a, b, a, b) (a, b, b, a)
(b, b, a, a) (b, a, b, a) (b, a, a, b)

6 choix

en tout $\frac{n(n-1)}{2} \times 6 = 3n(n-1)$

Donc

$$\mathbb{E}\left[\left(\frac{X_1+\cdots+X_n}{n}\right)^4\right] \leq \frac{C}{n^2}$$

d'où le résultat.

Corollaire Si $(A_n)_{n\geq 1}$ est une suite d'événements ⫫ de même probabilité, on a

$$\underbrace{\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{A_k}}_{\text{fréquence de réalisation}} \xrightarrow[n\to\infty]{p.s} \mathbb{P}(A_1)$$

Ce corollaire fait le lien entre notre approche axiomatique moderne des probabilités et la définition historique de la probabilité comme fréquence d'apparition d'un événement quand on répète un grand nombre de fois une expérience aléatoire