Intégration et probabilités

Cours n° : 18

25/04/2024

$\mathbb{E}[X]$ espérance de X

$\mathbb{E}[f(X)]$ — — la v.a. $f(X)$

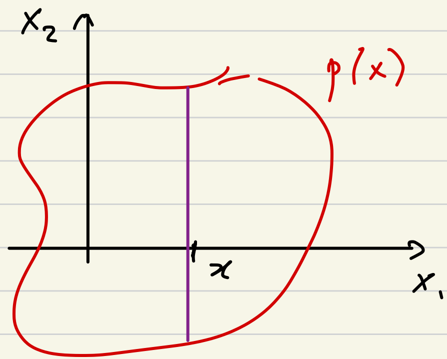
Si $\mathbb{E}[f(X)] = \int f \, d\nu$ pour "une grande classe de fonctions", alors la loi $P_X = \nu$.
$= \int f \, dP_X$

Ex : extraire les lois de v.a. individuelles X_j, à partir de la la loi globale du vecteur aléatoire $X = (X_1, .., X_d)$.

<u>Prop :</u> Soit $X = (X_1, .., X_d)$ une v.a. à valeurs dans $\mathbb{R}^d$. On suppose que X est une variable <u>à densité</u> $p(x_1, .., x_d)$.
Alors, $\forall j = 1, .., d$, la loi de la v.a. X_j est aussi à densité, donnée par :

$$\forall x \in \mathbb{R}, \quad p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, x_2 .. x_{j-1}, x, x_{j+1}, .., x_d) \, dx_1 .. dx_{j-1} \, dx_{j+1} .. dx_d$$

Les lois $P_{X_j} = p_j \, dx$ sont appelées les lois marginales de X.



$$p_1(x) = \int p(x, x_2)\, dx_2$$

On intègre sur la tranche verticale au-dessus de $x_1 = x$.

Preuve : $\pi_j : \mathbb{R}^d \to \mathbb{R}$, $(x_1, \dots, x_d) \mapsto x_j$. $\forall f : \mathbb{R} \to \mathbb{R}_+$ borélienne, le thme de Fubini-Tonnelli montre que :

$$\mathbb{E}[f(X_j)] = \mathbb{E}[f \circ \pi_j(X)] = \int f(x_j)\, p(x_1, \dots, x_d)\, dx_1 \dots dx_d$$

$$\overset{FT}{=} \int_{\mathbb{R}} f(x_j) \left(\int_{\mathbb{R}^{d-1}} p(x_1 \dots x_j, \dots x_d)\, dx_1 \dots dx_{j-1}\, dx_{j+1} \dots dx_d \right) dx_j$$

$$= \int_{\mathbb{R}} f(x_j)\, p_j(x_j)\, dx_j$$

$$\mathbb{E}[f(X_j)] = \int f(x)\, dP_{X_j}(x)$$

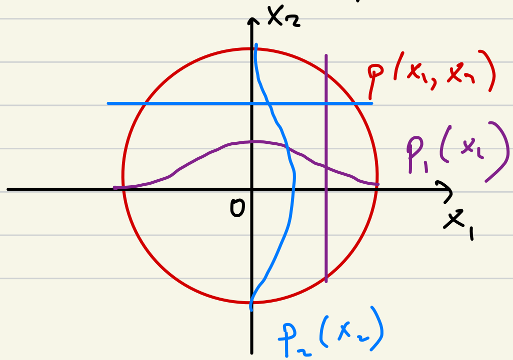
FT $\Rightarrow$ $p_j(x)$ est définie λ_1-p.p. est est λ_1-intégrable, d'intégrale $= 1$.

Comme $X_j = \pi_j(X) \Rightarrow$ la loi de X_j est déterminée par la loi de X par : $P_{X_j} = \pi_{j_*} P_X$.

Attention ! Connaître les lois individuelles $P_{X_1}, \dots, P_{X_d}$, ne permet pas de reconstruire la loi P_X.

En prenant les marginales de P_X, on perd de l'information.

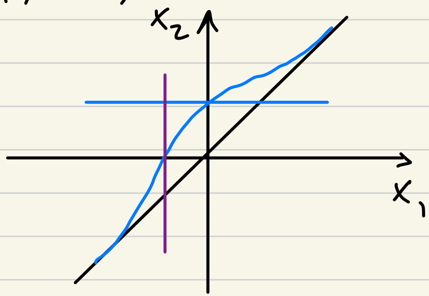
Ex : $X = (X_1, X_2)$, ayant une densité $p(x_1, x_2)$ à symétrie circulaire : $p(x_1, x_2) = \tilde{p}\left(\sqrt{x_1^2 + x_2^2}\right)$.



$$p_1(x_1) = \int p(x_1, x_2)\, dx_2$$

p_1 et p_2 sont égales à la même fonction $p : \mathbb{R} \to \mathbb{R}_+$.

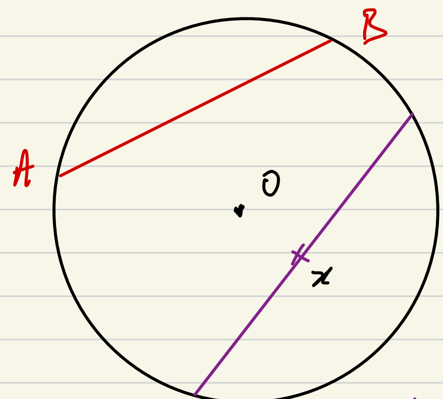
Il existe une autre loi P_X, donnant lieu aux mêmes marginales :
$X' = (X'_1, X'_2)$, dont la loi $P_{X'}$ est supportée par la diagonale $\{x_1 = x_2\}$



$$P_{X'} = p(x_1)\, \delta_{x_1 = x_2}$$

On vérifie que les deux marginales de $P_{X'}$ sont $p(x_1)\,dx_1$ et $p(x_2)\,dx_2$.

Exemple : expérience aléatoire : comment choisir au hasard une corde dans 1 cercle ?



Plusieurs façons de choisir la corde au hasard

a) choisir 2 pts A et B, et les relier sur le cercle

b) On choisit au hasard le milieu de la corde dans le disque $\{|x| \leq 1\}$ → 1 corde au hasard

Ces 2 manières de choisir 1 corde au hasard ne donnent pas les mêmes lois. On le montre en calculant la probabilité de l'évènement $\{|x| > 1/\sqrt{3}\}$.

$\Rightarrow$ ces 2 façons de choisir 1 corde au hasard ne sont pas équivalentes.

Exemples de lois "classiques"

a) Si $\# E = n \in \mathbb{N}^*$, une v. a. X est de loi uniforme sur E si $\forall x \in E$, $P(X = x) = \dfrac{1}{n}$.

Ex : tirage d'un dé ($n=6$)
tirage d'une pièce à "pile ou face" ($n = 2$)

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

c) Loi binomiale $\mathcal{B}(n,p)$, $n \in \mathbb{N}^*$, $p \in [0,1]$.

X est à valeurs dans $[\![0,n]\!]$, et

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \leq k \leq n.$$

Faire n tirages de Bernoulli-p, et ~~se~~ demander combien de fois on a obtenu $\{1\}$.

d) Loi géométrique de paramètre p : X à valeurs dans $\mathbb{N}$,

$$\mathbb{P}(X=k) = \frac{1}{1-p} p^k, \quad k \in \mathbb{N}.$$

Tirages de Bernoulli-p : nombre de $\{0\}$ avant le premier $\{1\}$

e) Loi de Poisson de paramètre $\lambda > 0$. X à valeurs dans $\mathbb{N}$,

$$\mathbb{P}(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}.$$

Décrit les apparitions d'évènements rares, pendant une période longue.

Suite (X_n) de v.a. réelles, de lois binomiales $\mathcal{B}(n, \frac{\lambda}{n})$

$$\Rightarrow \lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

$$\lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

Lois continues

v.a. continues, par ex. X à valeur dans $\mathbb{R}$.

a) loi uniforme sur $[a,b]$ $\to$ $p(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$

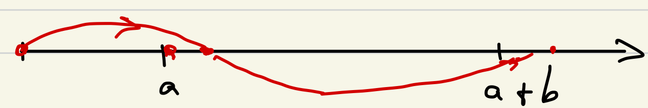
b) Loi exponentielle de paramètre $\lambda > 0$, supportée sur $\mathbb{R}_+$

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

Propriété suivante : $P(X > a+b) = P(X>a)\, P(X>b)$.

Processus sans mémoire

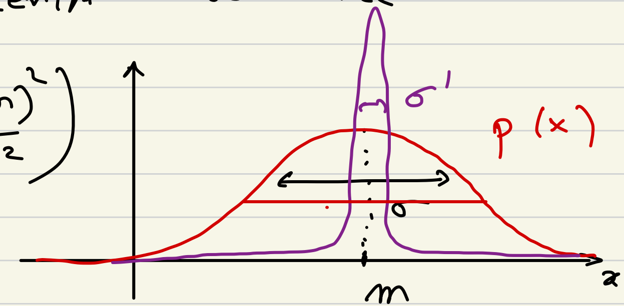
Les 2 sauts sont indépendants



c) Loi gaussienne, ou loi normale $\mathcal{N}(\underbrace{m}_{\text{centre}}, \underbrace{\sigma^2}_{\text{variance}})$ $\quad m \in \mathbb{R}$, $\sigma \geq 0$

X à valeur dans $\mathbb{R}$, de densité

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



Loi en forme de "cloche".

$m = \mathbb{E}[X]$

$\sigma^2 = \mathbb{E}[(X-m)^2]$ variance de X.

$\sigma = \sqrt{\sigma^2}$ l'écart-type

Si $m = m'$, mais $\sigma' \ll \sigma \Rightarrow$ la gaussienne $\mathcal{N}(m, \sigma'^2)$ est plus "piquée" que $\mathcal{N}(m, \sigma^2)$.

La v.a. $X - m$ suit la loi centrée $\mathcal{N}(0, \sigma^2)$

$\hookrightarrow \mathbb{E}[\cdot] = 0$

La v.a. $\dfrac{X-m}{\sigma}$ suit une loi centrée réduite $\mathcal{N}(0,1)$.

Si X est de loi $\mathcal{N}(m, \sigma^2) \Rightarrow \forall \lambda, \mu \in \mathbb{R}$, la v.a.

$\lambda X + \mu$ est de loi $\mathcal{N}(\lambda m + \mu, \lambda^2 \sigma^2)$.

Cette loi gaussienne apparaîtra de façon importante dans le théorème central limite. $\left(S_N = \dfrac{1}{\sqrt{N}} \displaystyle\sum_{j=1}^{N} X_j \right.$ ↘ v.a. indépendantes identiquement distribuées $\Big)$

- <u>Fonction de répartition d'1 v.a. réelle</u>

X v.a. réelle, $F_X : \mathbb{R} \to [0,1]$

$$t \mapsto F_X(t) = \mathbb{P}(X \leq t) = P_X(]-\infty, t])$$

F_X est croissante, continue à droite, $F_X(t) \underset{t \to -\infty}{\longrightarrow} 0$

$F_X(t) \underset{t \to +\infty}{\longrightarrow} 1$

Toute fonction F ayant ces propriétés correspond à une et une seule mesure de probabilité μ sur $\mathbb{R}$:

$$\exists !\ \mu \text{ mesure de proba sur } \mathbb{R} \text{ t.q. } F(t) = \mu(]-\infty, t]),\ \forall t.$$

points de discontinuité de F_X $\longleftrightarrow$ atomes de P_X.

- Tribu engendrée par une v.a.

X v.a. à image dans $(E, \mathcal{E})$. La tribu engendrée par X, notée $\sigma(X)$, est la + petite tribu sur Ω qui rende la fonction X mesurable.

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

- $(X_i)_{i \in I}$ famille de v.a. à valeurs dans $(E_i, \mathcal{E}_i)$.

$\longrightarrow$ tribu engendrée par $(X_i)_{i \in I}$:

$$\sigma((X_i)_{i \in I}) = \sigma\left(\bigcup_i \{ X_i^{-1}(B) ;\ B \in \mathcal{E}_i \} \right)$$

Prop°: Soit X une v.a. à valeurs dans $(E, \mathcal{E})$, et soit Y une v.a. réelle.

On a équivalence :

i) la v.a. Y est $\sigma(X)$-mesurable ($\Leftrightarrow \sigma(Y)$ est une sous-tribu de $\sigma(X)$)

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

Y est dite <u>totalement dépendante</u> de X.

Preuve : ii) $\to$ i) facile. $\forall A \subset \mathbb{R}$ borélien, $f^{-1}(A) \in \mathcal{E}$,

donc $X^{-1}(f^{-1}(A)) = Y^{-1}(A)$ est dans $\sigma(X)$.

i) $\Rightarrow$ ii) : supposons que Y est $\sigma(X)$-mesurable.

<u>1er cas</u> : si Y est étagée. $Y = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{A_i}$ (représent° canonique)

$\Rightarrow$ chaque $A_i \in \sigma(X)$.

$\Rightarrow \forall i, \exists B_i \in \mathcal{E}$ tq $A_i = X^{-1}(B_i) \Leftrightarrow \mathbb{1}_{A_i} = \mathbb{1}_{B_i} \circ X$.

B_i disjoints $\quad Y = \sum_i \lambda_i \mathbb{1}_{B_i} \circ X = f \circ X$, avec

$$f = \sum_i \lambda_i \mathbb{1}_{B_i} \quad \text{qui est mesurable } E, \mathcal{E} \to \mathbb{R}.$$

Cas général : $\quad Y = Y_+ - Y_-$

$\Rightarrow Y = \lim_{n \to \infty} Y_n$, où Y_n étagées, $\sigma(X)$-mesurables.

Chaque $Y_n = f_n \circ X$, pour $f_n : E \to \mathbb{R}$ mesurables.

$Y_n \to Y \Rightarrow f_n(x)$ converge, dès que x est dans l'image de X.

$$f(x) = \begin{cases} \lim f_n(x), & \text{si } x \in Im(X) \\ 0, & \text{sinon.} \end{cases}$$

f est mesurable : $\mathcal{E} \to \mathbb{R}$.

$\forall \omega \in \Omega$, on a $\underbrace{\lim_{n \to \infty} f_n(X(\omega))}_{f(X(\omega))} = \lim_{n \to \infty} Y_n(\omega) = Y(\omega)$.

$\square$

Comment tester une v.a. dont on ne connaît pas la loi ?

→ calculer les différents moments de X.

(X v.a. réelle). $p \in \mathbb{N}^*$. Le moment d'ordre p de la v.a. X est donné par $\mathbb{E}[X^p]$.

C'est bien défini si X une v.a. ≥ 0, ou si

$$\mathbb{E}[|X|^p] < \infty.$$

↳ $X \in \mathcal{L}^p(\Omega, \mathbb{P})$.

Rem : le moment d'ordre 1 est égal à l'espérance de X.

Déf : Une v.a. X est dite centrée si elle est intégrable et $\mathbb{E}[X] = 0$.

Réécrire le TCM : $X_n \geq 0$, $X_n \uparrow X \Rightarrow \mathbb{E}[X_n] \xrightarrow[n\to\infty]{} \mathbb{E}[X]$.

Lemme de Fatou : $X_n \geq 0 \rightarrow \mathbb{E}\left[\liminf_n X_n\right] \leq \liminf \mathbb{E}[X_n]$

TCD : $|X_n| \leq Z$ avec $\mathbb{E}[Z] < \infty$, $X_n \longrightarrow X$ p.p.

$$\Rightarrow \mathbb{E}[X_n] \xrightarrow[n\to\infty]{} \mathbb{E}[X]$$

- $L^p(\Omega, \mathcal{A}, \mathbb{P})$ Hölder :

$$\mathbb{E}[|XY|] \leq \left(\mathbb{E}[|X|^p]\right)^{\frac{1}{p}} \left(\mathbb{E}[|Y|^q]\right)^{\frac{1}{q}}$$

$p, q \in [1, \infty]$, conjugués

$\mathbb{P}$ mesure de proba $\Rightarrow$ $\|X\|_1 \leq \|X\|_p$, $\forall p \geq 1$

$\|X\|_r \leq \|X\|_p$ $\forall p \geq r \geq 1$

ex : Cauchy - Schwarz :

$$\mathbb{E}[|XY|] \leq \left(\mathbb{E}[|X|^2]\right)^{\frac{1}{2}} \left(\mathbb{E}[Y^2]\right)^{\frac{1}{2}}$$

$$\Rightarrow \mathbb{E}[|X|]^2 \leq \mathbb{E}[X^2].$$

Déf°: Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs réelles.

La __variance__ de X est définie par

$$\mathrm{var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \geq 0$$

Son Ecart-type $\sigma_X := \sqrt{\mathrm{var}(X)}$

Si $\mathrm{var}(X) = 0 \Rightarrow X = \mathbb{E}[X]$ $\underbrace{\text{p.s.}}_{\text{presque sûrement}}$

$= \mathbb{P}$-p.p.

Ex. de la loi normale $\mathcal{N}(m, \sigma^2) =: \rho_X$: variance

$$\mathrm{var}(X) = \sigma^2.$$

σ représente la "largeur à mi-hauteur". $p(x)$ la densité de X

$$\frac{p(x)}{p_{max}} \geq \frac{1}{2} \iff |x - m| \leq \sigma\sqrt{2\log 2}$$

$$\frac{p(x)}{p_{max}} \geqslant e^{-\frac{1}{2}} \iff |x-m| \leq \sigma$$

Prop : Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. Sa variance peut s'exprimer par :

$$\operatorname{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$\forall a \in \mathbb{R}$, on a : $\mathbb{E}[(X-a)^2] = \operatorname{var}(X) + (\mathbb{E}[X]-a)^2$

$\Longrightarrow$ expression variationnelle

$$\operatorname{var}(X) = \inf_{a \in \mathbb{R}} \mathbb{E}[(X-a)^2].$$

le minimum est atteint en $a = \mathbb{E}[X]$.

Preuve : calcul facile.

. **Inégalité de Markov**

Soit X une v.a. positive.

$$\forall a > 0, \quad \text{on a} \quad \mathbb{P}(X \geq a) \leq \frac{1}{a} \mathbb{E}[X].$$

(intéressant si $\mathbb{E}[X] < \infty$)

. **Inégalité de Bienaymé-Tchebicheff**

Soit X une v.a. dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Alors, $\forall a > 0$,

$$\mathbb{P}(\underbrace{|X - \mathbb{E}[X]| \geq a}) \leq \frac{1}{a^2} \operatorname{var}(X).$$

écart de X p/r à sa moyenne

$\simeq$ dispersion de la v.a. X.

Preuve : utiliser Markov pour la v.a. $Y = (X - \mathbb{E}(X))^2$.

Plusieurs v.a. réelles $\rightsquigarrow$ notion de <u>covariance</u>, pour mesurer les <u>corrélations</u> entre les différentes v.a.

<u>Déf°</u> Soit $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. La covariance de X et Y est définie par :

$$\mathrm{cov}(X,Y) = \mathbb{E}\Big[\underbrace{(X - \mathbb{E}[X])}_{\text{v.a. centrée}}\,(Y - \mathbb{E}[Y])\Big] \in \mathbb{R}$$

$= \mathrm{cov}(Y, X)$

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

$$= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Plus généralement, pour $X = (X_1, \dots, X_d)$ à valeurs dans $\mathbb{R}^d$, avec $X_j \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, la <u>matrice de covariance</u> est définie par

$$K_X = \big(\mathrm{cov}(X_i, X_j)\big)_{i,j=1,\dots,d}$$

Propriétés :

1) $\mathrm{cov}(X,X) = \mathrm{var}(X) \geq 0.$

Par contre, $\mathrm{cov}(X,Y)$ peut prendre des valeurs ≥ 0 ou ≤ 0.

ex : $Y = -X$

$\Rightarrow \mathrm{cov}(X,Y) = -\mathrm{var}(X).$

2) $|\mathrm{cov}(X,Y)| \leq \sqrt{\mathrm{var}(X)\ \mathrm{var}(Y)}$

⚠ homogénéité des expressions

3) $(X,Y) \longmapsto \mathrm{cov}(X,Y)$ est une forme bilinéaire sur $L^2(\Omega, \mathbb{P})$.

4) $X = (X_1, \dots, X_d)$ avec $X_j \in L^2 \Rightarrow K_X$ est une matrice symétrique positive : $\forall \lambda = (\lambda_1, \dots, \lambda_d)$,

$$\langle \lambda, K_X \lambda \rangle = \sum_{i,j=1}^{d} \lambda_i \lambda_j K_X(i,j) = \sum_{i,j=1}^{d} \lambda_i \lambda_j \mathbb{E}\left((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}[X_j]) \right)$$

$$= \mathrm{var}\left(\sum_{i=1}^{d} \lambda_i X_i \right) \geq 0$$

Ex : Si A matrice $n \times d$, et $X' = AX$ (X' est un n-vecteur aléatoire)

$$\Rightarrow K_{X'} = A K_X {}^tA$$ matrice $n \times n$ symétrique ≥ 0.

Rem : si X, Y sont des v. a. centrées

Si $\mathrm{cov}(X, Y) > 0$ indique que, lorsque $X(\omega)$ est ≥ 0, il y a de fortes chances que $Y(\omega)$ soit aussi ≥ 0. et inversement.

Si $\mathrm{cov}(X, Y) < 0$ ... il y a de fortes chances que $Y(\omega) < 0$.

. Si X et Y sont <u>indépendantes</u>, alors $\mathrm{cov}(X, Y) = 0$.

La réciproque est fausse !

- Régression linéaire : (v.a. réelles dans $L^2(\Omega, \mathbb{P})$)

Famille de v-a $(Y_1, \dots, Y_n)$ " de référence ".

Autre v.a. X " inconnue ".

Objectif : approximer X par une combinaison affine des $(Y_1, \dots, Y_n)$: de la forme

$$\beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \dots + \beta_n Y_n$$

$\Rightarrow$ On cherche à <u>minimiser</u> l'espérance suivante

$$\mathbb{E}\left[\left(X - (\beta_0 + \beta_1 Y_1 + \dots + \beta_n Y_n) \right)^2 \right]$$

$$\left(\mathbb{E}\, | X - (\beta_0 + \dots + \beta_n Y_n) | \leftarrow + \text{ difficile à calculer.} \right)$$

par rapport au choix des coefficients $\beta_0, \dots, \beta_n$.

Problème de minimisation, qu'on appelle <u>régression linéaire</u>.

<u>Prop°</u> Ce problème de minimisation admet une unique solution : $\inf_{(\beta_i) \in \mathbb{R}^{n+1}} \mathbb{E}[(\qquad)^2] = \mathbb{E}[(X - Z)^2]$, où

la v.a. Z est donnée par

$$Z = \mathbb{E}[X] + \sum_{i=1}^{d} \alpha_i (Y_i - \mathbb{E}[Y_i]) \quad , \text{ où}$$

les coefficients α_i sont solutions du système linéaire :

$$\sum_{i=1}^{n} \alpha_i \underbrace{\mathrm{cov}(Y_i, Y_k)}_{K_Y(i,k)} = \mathrm{cov}(X, Y_k), \quad \forall k = 1, \dots, n.$$

Si K_Y est inversible $\Rightarrow \exists!$ solution $(\alpha_i)_{i=1,\dots n}$,

$$\alpha = \mathrm{cov}(X,Y)\, K_Y^{-1}.$$

Prop : 5.15

Hyp : $(E, \mathcal{A}, \mu)$ avec $\mu(E) < \infty$.

$r > 1$, $f_n \in L^r(\mu)$, $\int |f_n|^r \, d\mu \leq C \quad \forall n$.

$f_n(x) \longrightarrow f(x)$, $\forall x \in E$

Alors $f_n \xrightarrow{L^p} f$ pour tout $1 \leq p < r$.

Preuve : 1. Montrer que $f \in L^r(E, \mu)$.

Fatou : $\int |f|^r \, d\mu = \int (\underline{\lim} \, |f_n|^r) \, d\mu \leq \underline{\lim} \int |f_n|^r \, d\mu \leq C$.

2. On choisit $\varepsilon > 0$. $\forall n$, on considère l'ensemble

$$A_{n,\varepsilon} = \{ x \in E ; \ |f_n(x) - f(x)| \leq \varepsilon \}$$

Alors, l'hypothèse $f_n(x) \to f(x)$ implique que

$\forall x \in E, \exists N \geq 0, \forall n \geq N, \quad x \in A_{n,\epsilon}$. Donc $x \in \widehat{A}_{N,\epsilon} = \bigcap_{n \geq N} A_{n,\epsilon}$.

Donc $\bigcup_{N \geq 0} \widehat{A}_{N,\epsilon} = E$. Les $(\widehat{A}_{N,\epsilon})_N$ forment une suite $\uparrow$

Soit $\eta > 0$. Alors, $\exists N = N(\epsilon, \eta)$, t.q $\mu(\widehat{A}_{N,\epsilon}) \geq \mu(E) - \eta$.

$$\Leftrightarrow \mu(\widehat{A}_{N,\epsilon}^c) \leq \eta.$$

On veut borner $\int |f_n - f|^p \, d\mu$, pour $p \in [1, r[$.

On décompose $\int |f_n - f|^p \, d\mu = \int |f_n - f|^p \mathbb{1}_{\widehat{A}_{N,\epsilon}} \, d\mu + \int |f_n - f|^p \mathbb{1}_{\widehat{A}_{N,\epsilon}^c} \, d\mu$.

Le 1^{er} terme est borné par $\int |f_n - f|^p \mathbb{1}_{\widehat{A}_{N,\epsilon}} \, d\mu \leq \epsilon^p \mu(\widehat{A}_{N,\epsilon}) \leq \epsilon^p \mu(E)$.

Le 2^e terme est borné en utilisant Hölder :

$$\int |f_n - f|^p \mathbb{1}_{\widehat{A}_{N,\epsilon}^c} \, d\mu \leq \left(\int |f_n - f|^{p \frac{r}{p}} \, d\mu \right)^{\frac{p}{r}} \left(\int \mathbb{1}_{\widehat{A}_{N,\epsilon}^c} \, d\mu \right)^{1 - \frac{p}{r}}$$

$$\leq \|f_n - f\|_r^p \quad \mu(\widehat{A}_{N,\epsilon}^c)^{\frac{r-p}{r}} \leq (2C^{\frac{1}{r}})^p \quad \eta^{\frac{r-p}{r}}$$

En choisissant ε et η petits, et $N \geq N(\varepsilon,\eta)$, on a alors, $\forall n \geq N$,

$$\int |f_n - f|^p \, d\mu \leq \underbrace{\varepsilon^p \, \mu(E) + (2C^{\frac{1}{r}})^p \, \eta^{\frac{r-p}{r}}}_{\text{arbitrairement petit.}}$$

Conclusion : $\displaystyle\int |f_n - f|^p \, d\mu \xrightarrow[n \to \infty]{} 0$. $\square$