Espérance conditionnelle (chapitre 11 du cours de Le Gall)

Si A et B sont 2 événements d'un espace de probabilités $(\Omega, \mathcal{F}, P)$, si $P(B) > 0$, alors $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

"probabilité de A sachant B"
"A conditionnellement à B"

Plus généralement, si B est un événement, $P(B) > 0$,

$$\begin{aligned} \mathcal{F} &\xrightarrow{\mathbb{P}_B} \mathbb{R} \\ A &\longmapsto P(A|B) \end{aligned}$$

alors $\mathbb{P}_B$ est une mesure de probabilités

* $\mathbb{P}_B(\emptyset) = 0 = \dfrac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = 0$
* $\mathbb{P}_B(\Omega) = 1 = \dfrac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(B)}{\mathbb{P}(B)}$
* si $(A_n)_{n \in \mathbb{N}}$ sont des événements deux à deux disjoints :

$\forall m \neq n$, $A_m \cap A_n = \emptyset$, alors
$$\mathbb{P}_B\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) = \sum_{n \in \mathbb{N}} \mathbb{P}_B(A_n)$$

$$\mathbb{P}_B\Big(\bigcup_n A_n\Big) = \frac{\mathbb{P}\left(\left(\bigcup_n A_n\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_n (A_n \cap B)\right)}{\mathbb{P}(B)} = \sum_n \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)}$$
$$= \sum_n \mathbb{P}_B(A_n)$$

Cas discret : X v.a. à valeurs dans $A \subset \mathbb{R}^d$, A dénombrable (exple $A = \mathbb{N}$, $A = \mathbb{Z}$)

on peut définir une nouvelle variable aléatoire X_B (B, $\mathbb{P}(B) > 0$) par : $a \in A$

$$\mathbb{P}(X_B = a) = \mathbb{P}(X = a \mid B) = \frac{\mathbb{P}(\{X = a\} \cap B)}{\mathbb{P}(B)}$$

loi de X_B : loi conditionnelle de X sachant B

Espérance conditionnelle : Espérance de la loi conditionnelle.

Si $X \in L^1$, alors $X_B \in L^1$

$$\mathbb{E}(X_B) = \mathbb{E}(X|B) = \sum_{a \in A} \mathbb{P}(X_B = a)\, a = \sum_{a \in A} \frac{\mathbb{P}(\{X=a\} \cap B)}{\mathbb{P}(B)}\, a$$

bien défini car $\displaystyle\sum_{a \in A} \frac{\mathbb{P}(\{X=a\} \cap B)}{\mathbb{P}(B)} \|a\| \leq \frac{1}{\mathbb{P}(B)} \underbrace{\sum_{a \in A} \mathbb{P}(X=a) \|a\|}_{\text{bien défini}}$

remarque $\mathbb{E}(X|B) = \mathbb{E}(X_B) = \dfrac{\mathbb{E}(X \mathbb{1}_B)}{\mathbb{P}(B)}$

Espérance conditionnelle sachant $B \in \mathbb{R}^d$.

<u>Espérance conditionnelle sachant une variable aléatoire Y</u> (cas discret)

X, Y v.a à valeurs dans $A \subset \mathbb{R}^d$, A dénombrable ($\mathbb{N}^d$, $\mathbb{Z}^d$)

$$\mathbb{E}(X|Y) = \sum_{b \in A'} \mathbb{1}_{\{Y=b\}} \underbrace{\mathbb{E}_{\{Y=b\}}}(X) = \sum_{b \in A'} \mathbb{1}_{\{Y=b\}} \frac{\sum_{a \in A} \mathbb{P}(Y=b, X=a)\, a}{\mathbb{P}(Y=b)}$$

$A' = \{b \in A, \mathbb{P}(Y=b) > 0\}$

Si on pose $\varphi(y) = \begin{cases} \mathbb{E}_{\{Y=y\}}(X) & \text{pour } y \in A' \\ 0 & \text{pour } y \notin A' \end{cases}$

$$\mathbb{E}(X|Y) = \sum_{b \in A'} \mathbb{1}_{\{Y=b\}} \varphi(y) = \sum_{b \in A} \underbrace{\mathbb{1}_{\{Y=b\}} \varphi(y)}$$

variable aléatoire mesurable par rapport à la tribu $\sigma(Y)$ à valeurs dans $\mathbb{R}^d$

<u>Exemple</u> : on joue à pile ou face 3 fois, $Y = \mathbb{1}_{\{\text{pile au 1er lancer}\}}$

$X =$ nombre de "pile"

$$\mathbb{E}(X \mid \{Y=1\}) = \mathbb{E}(\text{nb de "pile"} \mid \text{pile au 1}^{\text{er}} \text{ lancer})$$
$$= \underbrace{1} + \frac{1}{2} + \frac{1}{2} = 2$$

$$\mathbb{E}(X \mid \{Y=0\}) = \mathbb{E}(\text{nb de "pile"} \mid \text{face au 1}^{\text{er}} \text{ lancer})$$
$$= \underbrace{0} + \frac{1}{2} + \frac{1}{2} = 1$$

$$\mathbb{E}(X \mid Y) = Y + 1$$

Rappel : une variable aléatoire Z est mesurable par rapport $\sigma(Y)$ si et seulement si elle peut s'exprimer $Z = \varphi(Y)$ où φ est une fonction mesurable) déf de Z mesurable par rapport à $\sigma(Y)$:
$\forall$ C borélien $\{Z \in C\} \in \sigma(Y)$

Rappel : Espace de Hilbert : espace vectoriel normé complet où la norme issue d'un produit scalaire : $\exists$ ϕ bilinéaire symétrique

$$x, y \mapsto \phi(x,y) = \phi(y,x)$$

$$\phi(\lambda x + \mu x', y) = \lambda \phi(x,y) + \mu \phi(x',y)$$

$$\phi(x, \lambda y + \mu y') = \lambda \phi(x,y) + \mu \phi(x,y')$$

ϕ définie positive : $\underline{\phi(x,x) \geq 0}$, $\underline{\phi(x,x) = 0}$ si et seulement si $x = 0$

norme de x : $\sqrt{\phi(x,x)}$

<u>Exemple</u> E : {variables aléatoires dans $L^2(\mathbb{R})$, de moyenne nulle}

$\mathbb{E}(XY) = \phi(X,Y)$ $\quad$ ϕ symétrique bilinéaire $\quad$ ϕ positive

$\phi(X,X) = \mathbb{E}(X^2) \geq 0$ $\quad$ ϕ n'est pas définie positive

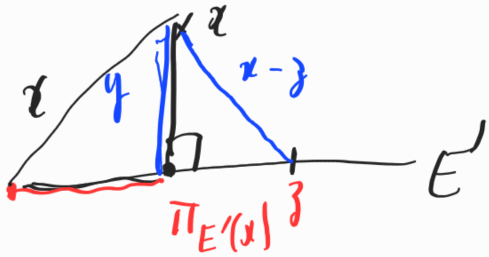
sur E, relation d'équivalence $\sim$: $X \sim Y$ si $X = Y$ ps

$(E/\sim)$ $\quad$ Ensemble quotient : ensemble des classes d'équivalence de E pour la relation $\sim$

Sur $(E/\sim)$ on définit ϕ comme plus haut [il faut vérifier $\phi(X,Y) = \phi(X',Y)$ si $X \sim X'$: $\phi(X,Y) - \phi(X',Y) = \phi(\underbrace{X-X'}_{0 \text{ p.s}}, Y) = 0$]

et ϕ est définie positive $\phi(X,X) = 0 \iff X = 0$ ps $\iff X \equiv 0$ dans $(E/\sim)$

Si E' est un sous-espace vectoriel d'un espace de Hilbert E, la projection orthogonale $\pi_{E'} : E \to E'$ est une application linéaire continue caractérisée par $x = \pi_{E'}(x) + y$ où y est orthogonal à E' : $\forall z \in E'$, $\phi(z, y) = 0$



$\pi_{E'}(x)$ minimise la fonction $z \mapsto \|x-z\|^2$ sur tous les $z \in E'$

$\|x-z\|^2 = \|y\|^2 + \|\pi_{E'}(x) - z\|^2 \geq \|y\|^2$

Sur l'espace E des v.a. $L^2(\mathbb{R})$, si $Y \in E$, alors l'ensemble E'_Y des variables aléatoires mesurables par rapport à $\sigma(Y)$ est un sous-espace vectoriel : $Z, Z' \in E'_Y$, $\lambda Z + \lambda' Z' \in E'_Y$

dans $\mathbb{R}^d$, E_d espace des variables aléatoires dans $L^2(\mathbb{R}^d)$

on définit $\phi(X,Y) = E(\langle X,Y\rangle)$ on généralise ce qui a été fait plus haut.

On va voir que si $X \in L^2(\mathbb{R}^d)$ à valeurs dans un ensemble dénombrable, si on note $m = \mathbb{E}(X)$, alors, $\mathbb{E}(X|Y) = m + Z$ où Z est la projection orthogonale de $(X-m)$ sur l'espace des variables aléatoires mesurable par rapport à $\sigma(Y)$

<u>Prop</u> : $A \subset \mathbb{R}$ dénombrable, $X \in L^2(\mathbb{R})$ à valeurs dans A, $m = \mathbb{E}(X)$

alors m minimise la fonction : $x \overset{f}{\longmapsto} \mathbb{E}[(X-x)^2]$

<u>dém</u> : $$f(x) = \sum_{i\in A} \mathbb{P}(X=i)\,(x-i)^2 = \sum_{i\in A} \mathbb{P}(X=i)(x^2-2ix+i^2)$$
$$= x^2 - 2x\,\mathbb{E}(X) + \mathbb{E}(X^2)$$

$f(x) \nearrow \infty$ $x \to +\infty$, $x \to -\infty$, $f'(x) = 2(x - \mathbb{E}(X))$ s'annule en $\mathbb{E}(X) = m$

on a un minimum en $x = m$

<u>Remarque</u> ds $\mathbb{R}^d$, $X \in L^2(\mathbb{R}^d)$, X à valeurs ds A dénombrable, $m = \mathbb{E}(X)$

m minimise la fonction $\mathbb{R}^d \longrightarrow \mathbb{R}_+$
$$x \longmapsto \mathbb{E}(\|X-x\|^2)$$
$$\sum_{i=1}^d (X_i - x_i)^2 \to \text{minimum en } m_i \text{ pour chaque } i$$

Prop : X, Y va à valeurs dans A dénombrable $\subset \mathbb{R}$ $X, Y \in L^2(\mathbb{R})$

Alors $\mathbb{E}(X|Y)$ minimise la fonction $Z \longmapsto \mathbb{E}((X-Z)^2)$ sur toutes les variables aléatoires Z qui sont $\sigma(Y)$-mesurables

dém $a \in A'$, conditionnellement à $\{Y=a\}$, $\mathbb{E}((X-x)^2|Y=a)$ est minimisée par le réel $\mathbb{E}_{\{Y=a\}}(X)$. En effet, $x \mapsto \mathbb{E}[(X-x)^2|Y=a]$ est $x \longmapsto \mathbb{E}_{\{Y=a\}}[(X-x)^2] = \mathbb{E}[(X_{\{Y=a\}}-x)^2]$ est minimisée par la moyenne de la variable aléatoire $X_{\{Y=a\}}$; cette moyenne est $\mathbb{E}(X_{\{Y=a\}})$

$$= \mathbb{E}_{\{Y=a\}}(X) = \sum_{b \in A'} \frac{\mathbb{P}(X=b, Y=a)}{\mathbb{P}(Y=a)} b$$

$\to$ vrai pour tout $a \in A'$

On cherche une variable aléatoire Z, $\sigma(Y)$-mesurable et qui minimise $\mathbb{E}(X-Z)^2 \Leftrightarrow$ on cherche une fonction φ qui minimise

$$\mathbb{E}[(X-\varphi(Y))^2] = \mathbb{E}\left[\sum_{a \in A'} \mathbb{1}_{\{Y=a\}}(X-\varphi(a))^2\right] = \sum_{a \in A'} \mathbb{E}(\mathbb{1}_{\{Y=a\}}(X-\varphi(a))^2)$$

φ doit minimiser $\mathbb{E}(\mathbb{1}_{\{Y=a\}}(X-\varphi(a))^2)$ pour tout $a \in A'$

$\to$ φ doit valoir $\mathbb{E}_{\{Y=a\}}(X)$ sur l'événement $\{Y=a\}$

$$\varphi(a) = \mathbb{E}_{\{Y=a\}}(X) \to \varphi(Y) = \mathbb{E}(X|Y)$$

→ on a minimisé $\underbrace{\ }_{\text{sur les variables aléatoires } \sigma(Y)\text{-mesurables}}$ $Z \longmapsto \mathbb{E}((X-Z)^2)$ en prenant $Z = \varphi(Y) = \mathbb{E}(X|Y)$

→ si $\mathbb{E}(X) = 0$, Z est bien la projection orthogonale de X sur le sous-espace vectoriel des variables aléatoires de moyenne nulle, $\sigma(Y)$-mesurables, de carré intégrable.

→ si $\mathbb{E}(X) = m$, par le raisonnement ci-dessus, la variable aléatoire Z $\sigma(Y)$-mesurable qui minimise $\mathbb{E}[((X-m)-Z)^2]$ est la projection orthogonale de $(X-m)$ sur le ss-espace vectoriel des variables aléatoires $\sigma(Y)$-mesurables et de carré intégrable ;

$$\mathbb{E}(((X-m)-Z)^2) = \mathbb{E}(X-(m+Z))^2 \rightarrow \mathbb{E}(X|Y) = m + Z$$

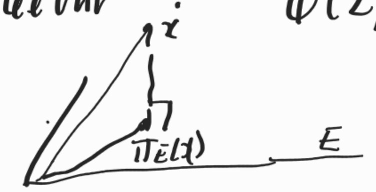
<u>Prop</u> cadre discret : v.a à valeurs dans A dénombrable $A \subset \mathbb{R}$, $X, Y \in L^1$

(1) $\mathbb{E}(|\mathbb{E}(X|Y)|) \leq \mathbb{E}(|X|)$

(2) $\forall Z$ v.a. $\sigma(Y)$-mesurable, $\mathbb{E}(ZX) = \mathbb{E}(Z\,\mathbb{E}(X|Y))$

remarque : si on se ramène à des v.a de moyenne nulle (on peut toujours le faire en remplaçant X par $X - \mathbb{E}(X)$) et si la v.a sont dans L^2,

(2) se réécrit : $\Phi(Z, X) = \Phi(Z, \mathbb{E}(X|Y))$



$y \in E$, $\Phi(y, x) = \Phi(y, \pi_E(x))$

<u>dém</u> (1) $\mathbb{E}|\mathbb{E}(X|Y)| = \sum_{a \in A'} \mathbb{P}(Y=a) \dfrac{|\mathbb{E}(X \mathbb{1}_{\{Y=a\}})|}{\mathbb{P}(Y=a)} \leq \sum_{a \in A} \mathbb{E}(|X| \mathbb{1}_{Y=a})$
$= \mathbb{E}|X|$

② si Z est $\sigma(Y)$-mesurable, $Z = \varphi(Y)$, φ mesurable

$$\mathbb{E}(Z\,\mathbb{E}(X|Y)) = \sum_{a\in A'} \varphi(a)\,\mathbb{E}(X\,\mathbb{1}_{(Y=a)}) = \sum_{a\in A'} \mathbb{E}(\varphi(Y)X\,\mathbb{1}_{\{Y=a\}})$$

$$= \mathbb{E}\Big(\sum_{a\in A'} \varphi(Y)\,X\,\mathbb{1}_{(Y=a)}\Big)$$

$$= \mathbb{E}\Big(\sum_{a\in A} \varphi(Y)\,X\,\mathbb{1}_{(Y=a)}\Big)$$

$$= \mathbb{E}(\varphi(Y)\,X)$$