Probabilités

Chapitre 8 : Espérance conditionnelle

Lucie Le Briquer

23 novembre 2017

1 Échauffement

- Si X est une v.a., la plus petite tribu incluse dans $\mathcal{A}$ telle que X soit mesurable est :

$$\sigma(X) := \{(X \in B) \mid B \in \mathcal{A}\}$$

- X est $\mathcal{B}-$mesurable si $X\colon (\Omega, \mathcal{B}) \to (E, \mathcal{E})$ est mesurable $\Leftrightarrow \sigma(X) \subseteq \mathcal{B}$
- X est $Y-$mesurable si X est $\sigma(Y)-$mesurable $\Leftrightarrow \sigma(X) \subseteq \sigma(Y)$

Lemme 1 (crucial)
Soit Y une v.a. à valeurs dans $(E, \mathcal{E})$, et X une v.a. réelle.

$$X \text{ est } Y - \text{mesurable} \quad \Leftrightarrow \quad \exists h\colon (E, \mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ mesurable telle que } X = h(Y)$$

$$\begin{tikzcd}
\Omega \arrow[rr, "Y"] \arrow[dr, "X"'] & & E \arrow[dl, dashed, "h"] \\
& \mathbb{R} &
\end{tikzcd}$$

Remarques.

- Ce résultat est de la théorie de la mesure, et ne dépend pas de $\mathbb{P}$.
- En appliquant coordonnées par coordonnées, le résultat est vrai pour $X \in \mathbb{R}^d$.
- Il permet de parler de mesurabilité sans parler de tribus.

Preuve.
$\Leftarrow$: Si $X = h(Y)$ alors :

$$\begin{aligned}
\sigma(X) &= \{(X \in B) \mid B \in \mathcal{B}(\mathbb{R})\} = \{h(Y) \in B \mid B \in \mathcal{B}(\mathbb{R})\} \\
&= \{Y \in h^{-1}(B)\} \subseteq \{Y \in C \mid \mathcal{E}\} = \sigma(Y)
\end{aligned}$$

$\Rightarrow$:

- Si $X = \mathbb{1}_A$, $A = X^{-1}(\{1\}) \in \sigma(X) \subset \sigma(Y)$. Donc $\exists C \in \mathcal{E}$ tel que $A = Y^{-1}(C)$. Posons alors :
$$h\colon \left\{ \begin{array}{ccc} E & \longrightarrow & \mathbb{R} \\ y & \longmapsto & \mathbb{1}_{y \in C} \end{array} \right.$$
Ainsi $h(Y) = \mathbb{1}_A = X$.

- Si $X = \sum_{i=1}^d \lambda_i \mathbb{1}_{A_i}$. On peut supposer les A_i disjoints puisque si $A_i \cap A_j \neq \emptyset$ alors $\lambda_i \mathbb{1}_{A_i} + \lambda_j \mathbb{1}_{A_j} = \lambda_i \mathbb{1}_{A_i \setminus A_j} + \lambda_j \mathbb{1}_{A_j \setminus A_i} + (\lambda_i + \lambda_j)\mathbb{1}_{A_i \cap A_j}$. On peut aussi supposer les λ_i deux à deux distincts car si $\lambda_i = \lambda_j$ alors $\lambda_i \mathbb{1}_{A_i} + \lambda_j \mathbb{1}_{A_j} = \lambda_i \mathbb{1}_{A_i \cup A_j}$.
Alors $A_i = X^{-1}(\{\lambda_i\}) \in \sigma(X) \subseteq \sigma(Y)$ donc $\exists C_i \in \mathcal{E}$ tel que $A_i = Y^{-1}(C_i)$ qu'on peut choisir disjoints.
$$h(y) := \sum_{i=1}^n \lambda_i \mathbb{1}_{y \in C_i}$$
Alors $\omega \in A_i \Rightarrow Y(\omega) \in C_i \Rightarrow h(Y(\omega)) = \lambda_i = X(\omega)$.

- Si X est positif alors $X\colon (\Omega, \sigma(Y)) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est limite croissante de fonctions étagées
$$X_n\colon (\Omega, \sigma(Y)) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$
$\exists h_n$ étagée tel que $X_n = h_n(Y)$. Posons $h(y) = \lim \uparrow h_n(y)$ si elle existe, 0 sinon. Alors :
$$X(\omega) = \lim \uparrow X_n(\omega) = \lim \uparrow h_n(Y(\omega)) = h(Y(\omega))$$

- Si X n'est pas positif, $X = X_+ - X_-$.

$\square$

2 Deuxième échauffement

Objectif. Définir $\mathbb{E}[X|Y]$ l'espérance conditionnelle de X sachant Y. Fonction de Y la plus proche de la v.a. X.

Exemple. Y discrète : $Y \subset \{y_i\}_{i \in I}$ avec I au plus dénombrable, $\mathbb{P}(Y = y_i) > 0$. X une v.a. réelle L^2. On cherche h tel que $\mathbb{E}[|h(Y) - X|^2]$ soit minimal. Si $Y\colon \Omega \to (E, \mathcal{E})$ et $h\colon E \to \mathbb{R}$:
$$\begin{aligned} \mathbb{E}\left[|h(Y) - X|^2 \left(\sum_{i \in I} \mathbb{1}_{Y = y_i} \right) \right] &= \sum_{i \in I} \mathbb{E}\left[\mathbb{1}_{Y = y_i} (h(y_i) - X)^2 \right] \\ &= \sum_{i \in I} \Big(h(y_i)^2 \mathbb{P}(Y = y_i) - 2h(y_i)\mathbb{E}[X\mathbb{1}_{Y = y_i}] + \mathbb{E}[X^2 \mathbb{1}_{Y = y_i}] \Big) \end{aligned}$$

On peut choisir chaque $h(y_i)$: il faut minimiser un polynôme de degré 2.
$$h(y_i) = \frac{\text{“} - b''}{2a} = \frac{\mathbb{E}[X\mathbb{1}_{Y = y_i}]}{\mathbb{P}(Y = y_i)}$$

Remarque. Conditionnement par rapport à un événement B tel que $\mathbb{P}(B) > 0$:

- On définit :
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$
Fait important : $A \longmapsto \mathbb{P}(A|B)$ est aussi une probabilité.

- On définit $\mathbb{E}[.|B]$ l'espérance sous $\mathbb{P}(.|B)$. On a :
$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbb{1}_B]}{\mathbb{P}(B)}$$
car
$$\mathbb{E}[\mathbb{1}_A|B] = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{E}[\mathbb{1}_A\mathbb{1}_B]}{\mathbb{P}(B)}$$
$+$ linéarité $+$ limite croissante de fonctions étagées.

Bilan. Un h tel que $\mathbb{E}[|X - h(Y)|^2]$ soit minimal :

$$h(y) := \begin{cases} \mathbb{E}[X|Y = y_i] & \text{si } y = y_i \\ 0 \quad \text{si } y \notin \{y_1, ..., y_n, ...\} \end{cases}$$

Dans ce cas on verra que $\mathbb{E}[X|Y] = h(Y)$. $\mathbb{E}[X|Y]$ fait une moyenne de X sur le lieu où Y a une valeur fixée.

Remarque. $h(Y)$ ainsi défini a un sens dès que $X \in L^1$. On va vouloir définir $\mathbb{E}[X|Y]$ $\forall X \in L^1$.

3 Définition, construction

Propriété 1

Soit $\mathcal{B}$ une tribu et X une v.a. réelle L^1. Alors il existe une unique ($=$ à un ensemble négligeable près) v.a. réelle Z telle que :

$$\begin{cases} Z \text{ est } \mathcal{B}\text{-mesurable} \\ \forall W \text{ v.a. réelle } \mathcal{B}\text{-mesurable bornée, } \mathbb{E}[XW] = \mathbb{E}[ZW] \quad \text{(propriété caractéristique)} \end{cases}$$

Définition 1 (espérance conditionnelle)

On appelle le Z de la propriété précédente l'*espérance conditionnelle de X sachant $\mathcal{B}$* et on le note $\mathbb{E}[X|\mathcal{B}]$. On note aussi $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$.

Remarque. Intuition : à côté d'une v.a. $\mathcal{B}-$mesurable, mettre X ou Z donne le même résultat. Donc Z est proche de X parmi les $\mathcal{B}-$mesurables.

Preuve. (de la propostion)

Unicité. Si Z_1, Z_2 vérifient la propriété caractéristique, ils sont $\mathcal{B}-$mesurables. Donc $W = \mathbb{1}_{Z_1 > Z_2}$ aussi et alors :

$$\left\{ \begin{array}{l} \mathbb{E}[Z_1 W] = \mathbb{E}[XW] \\ \mathbb{E}[Z_2 W] = \mathbb{E}[XW] \end{array} \right. \Rightarrow \mathbb{E}[Z_1 W] = \mathbb{E}[Z_2 W]$$

En effectuant la différence :

$$\mathbb{E}[(Z_1 - Z_2)\mathbb{1}_{Z_1 > Z_2}] = 0$$

Donc $(Z_1 - Z_2)\mathbb{1}_{Z_1 > Z_2} = 0$ p.s. i.e. $Z_1 \leqslant Z_2$ p.s. Par symétrie on a de même $Z_2 \leqslant Z_1$ p.s. et donc $Z_1 = Z_2$ p.s.

Existence. Si $X \in L^2$. $B \subseteq \mathcal{A}$.

$$\mathcal{L}^2(\mathcal{A}) = \{X \text{ v.a.} \mid \mathbb{E}[|X|^2] < +\infty\}_{/(X \sim X' \text{ si } X = X' \text{ p.s.})}$$

a une structure d'espace de Hilbert avec $\langle X, Y \rangle = \mathbb{E}[XY]$. $\mathcal{L}^2(B)$ sev fermé de $\mathcal{L}^2(\mathcal{A})$ donc $\mathcal{L}^2(B) \oplus \mathcal{L}^2(B)^\perp = \mathcal{L}^2(\mathcal{A})$. Soit Π la projection orthogonale sur $\mathcal{L}^2(B)$ et $Z = \Pi X$.

- $Z \in \mathcal{L}^2(B)$ donc est $B-$mesurable.
- Si W est $B-$mesurable bornée, $W \in \mathcal{L}^2(B)$. Et comme $Z - X = \Pi X - X \in \mathcal{L}^2(B)^\perp$ on a :

$$\langle Z - X, W \rangle = 0 \quad \text{i.e.} \quad \langle Z, W \rangle = \langle X, W \rangle \quad \text{i.e.} \quad \mathbb{E}[ZW] = \mathbb{E}[XW]$$

$\square$

Remarques.

1. Sur L^2, $\mathbb{E}[.|B]$ est la projection orthogonale sur $\mathcal{L}^2(B)$.
2. $\mathbb{E}[.|B]$ est linéaire.
3. De plus, si $X \in L^1$ et $X \geqslant 0$ p.s. alors $\mathbb{E}[X|B] \geqslant 0$ p.s. En effet :

$$W := \mathbb{1}_{\mathbb{E}[X|B] < 0} \text{ est } B - \text{mesurable bornée}$$

Donc :

$$\mathbb{E}[\underbrace{W\mathbb{E}[X|B]}_{\leqslant 0}] = \mathbb{E}[\underbrace{WX}_{\geqslant 0}] \geqslant 0$$

Donc $\mathbb{1}_{\mathbb{E}[X|B] < 0}\mathbb{E}[X|B] = 0$ p.s. i.e. $\mathbb{E}[X|B] \geqslant 0$ p.s.

4. Croissance : si X et Y sont dans $\mathcal{L}^2$ alors :

$$X \leqslant Y \text{ p.s.} \Rightarrow \mathbb{E}[X|B] \leqslant \mathbb{E}[Y|B]$$

par linéarité et positivité.

Preuve. (existence si $X \geqslant 0$ p.s.)

Si $X \geqslant 0$ p.s., pour $n \in \mathbb{N}$ soit $X_n = \min(X_n) \in L^2$. Donc $\mathbb{E}[X_n|B]$ existe. (X_n) est une suite croissante de v.a. donc $(\mathbb{E}[X_n|B])$ aussi d'après les remarques précédentes.

Notons alors :

$$Z := \lim_{n \to +\infty} \uparrow \mathbb{E}[X_n|B] \quad \text{qui est } \mathcal{B} - \text{mesurable}$$

Si $W = \mathbb{1}_B$ avec $B \in \mathcal{B}$:

$$\begin{aligned}
\mathbb{E}[XW] &= \mathbb{E}\Big[(\lim \uparrow \min(X,n))\, W \Big] \\
&\overset{\text{TCM}}{=} \lim \uparrow \mathbb{E}[\min(X,n)W] \\
&\overset{\text{prop car}}{=} \lim \uparrow \mathbb{E}[\mathbb{E}[X_n|B]W] \\
&\overset{\text{TCM}}{=} \mathbb{E}\Big[\lim \uparrow \mathbb{E}[X_n|B]W \Big] \\
&= \mathbb{E}[ZW]
\end{aligned}$$

Pour $X \geqslant 0$ on a donc construit Z tel que Z est $\mathcal{B}-$mesurable et $\forall W$ $\mathcal{B}-$mesurable $\geqslant 0$ $\mathbb{E}[XW] = \mathbb{E}[ZW]$. $\rightarrow$ définition pour $X \geqslant 0$ p.s. $\square$

Preuve. (existence dans le cas L^1)

Si $X \in L^1$, $X = X^+ - X^-$ où X^+ et $X^- \geqslant 0$ p.s. On peut donc construire $Z_+ := \mathbb{E}[X^+|\mathcal{B}]$ et $Z_- := \mathbb{E}[X^-|\mathcal{B}]$. Posons $Z = Z_+ - Z_-$ $\mathcal{B}-$mesurable.

Vérifions que $\mathbb{E}[0|\mathcal{B}] = 0$:

$$0 \text{ est } \mathcal{B} - \text{mesurable et si } W \geqslant 0,\ \mathbb{E}[0W] = 0 = \mathbb{E}[\underbrace{\mathbb{E}[0|\mathcal{B}]W}_{\geqslant 0}]$$

Donc $\mathbb{E}[0|\mathcal{B}]W = 0$ $\forall W \geqslant 0 \Rightarrow \mathbb{E}[0,\mathcal{B}] = 0$.

Remarque. Si $X \geqslant 0$ et L^1, $X^- = 0$ p.s. donc $Z = Z^+ - 0 = Z^+ = \mathbb{E}[X|\mathcal{B}]$.

Retour à la démonstration. Par construction, $Z_+, Z_- \geqslant 0$ p.s. Par la propriété caractéristique avec $W = 1$:

$$\mathbb{E}[Z_+] = \mathbb{E}[X^+] < +\infty \quad \text{et} \quad \mathbb{E}[Z_-] = \mathbb{E}[X^-] < +\infty$$

Donc $Z_+, Z_- \in L^1$. Donc $Z \in L^1$. Si W est mesurable et bornée :

$$\begin{aligned}
\mathbb{E}[XW] &= \mathbb{E}[X^+W^+] + \mathbb{E}[X^+W^-] + \mathbb{E}[X^-W^+] + \mathbb{E}[X^-W^-] \\
&\underset{\text{prop car}}{=} \mathbb{E}[Z^+W^+] + ... \\
&= \mathbb{E}[ZW]
\end{aligned}$$

$\square$

Remarque. $\mathbb{E}[.]$ définie :

- sur les v.a. $\geqslant 0$

$$\mathbb{E} \colon \{\text{v.a.} \geqslant 0\} \longrightarrow \mathbb{R}^+ \cup \{+\infty\}$$

- sur les v.a. L^1 (t.q. $\mathbb{E}[|X|] < +\infty$))

$$\mathbb{E} \colon L^1 \longrightarrow \mathbb{R}$$

et $\mathbb{E}[.|\mathcal{B}]$ définie :

- sur les v.a. $\geqslant 0$:

$$\mathbb{E}[.|\mathcal{B}] \colon \{\text{v.a. } \geqslant 0\} \longrightarrow \{\text{v.a. } \mathcal{B}\text{-mesurables bornées}\}$$

  Caractérisée par :

$$\begin{cases} \mathbb{E}[X|\mathcal{B}] \text{ est } \mathcal{B}\text{-mesurable} \geqslant 0 \\ \forall W \ \mathcal{B}\text{-mesurable} \geqslant 0, \ \mathbb{E}[\mathbb{E}[X|\mathcal{B}]X] = \mathbb{E}[XW] \end{cases}$$

- sur les v.a. L^1 :

$$\mathbb{E}[.|\mathcal{B}] \colon \{\text{v.a. } \in L^1\} \longrightarrow \{\text{v.a. } \mathcal{B}\text{-mesurables } L^1\}$$

  Caractérisée par :

$$\begin{cases} \mathbb{E}[X|\mathcal{B}] \text{ est } \mathcal{B}\text{-mesurable} \\ \forall W \ \mathcal{B}\text{-mesurable bornée}, \ \mathbb{E}[\mathbb{E}[X|\mathcal{B}]X] = \mathbb{E}[XW] \end{cases}$$

Remarque. Formulation équivalente pour $X \in L^1$:

$$\forall W, \ \mathcal{B}\text{-mesurable bornée } \mathbb{E}[XW] = \mathbb{E}[ZW] \quad \Leftrightarrow \quad \forall B \in \mathcal{B}, \ \mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[Z\mathbb{1}_B]$$

Remarque. (espérance conditionnelle par rapport à une v.a. : $\mathcal{B} = \sigma(Y)$)

$$\begin{cases} Z \ Y\text{-mesurable} \\ \forall W \ Y\text{-mes. bornée } \mathbb{E}[WZ] = \mathbb{E}[YZ] \end{cases} \Leftrightarrow \begin{cases} \mathbb{E}[X|Y] = h(Y) \text{ caractérisée par} \\ \forall \varphi \text{ mes. bornée}, \ \mathbb{E}[X\varphi(Y)] = \mathbb{E}[h(Y)\varphi(Y)] \end{cases}$$

Exemple. Soit Y v.a. discrète, $Y \in \{y_i\}_{i \in I}$ p.s. tel que $\mathbb{P}(Y = y_i) > 0$ et I au plus dénombrable. Que vaut $\mathbb{E}[X|Y]$?
$\forall \varphi$ mesurable bornée,

$$\begin{aligned} \mathbb{E}[X\varphi(Y)] &= \mathbb{E}\left[X \sum_{i \in I} \varphi(y_i)\mathbb{1}_{Y = y_i}\right] \\ &= \sum_{i \in I} \varphi(y_i) \underbrace{\frac{\mathbb{E}[X\mathbb{1}_{Y=y_i}]}{\mathbb{P}(Y = y_i)}}_{h(y_i)} \mathbb{P}(Y = y_i) \qquad \text{(par Fubini car } \varphi \text{ bornée)} \\ &= \sum_{i \in I} \varphi(y_i)h(y_i)\mathbb{P}(Y = y_i) \\ &= \mathbb{E}[h(Y)\varphi(Y)] \end{aligned}$$

où $h(y) = \mathbb{E}[X|Y = y_i]$ si $y = y_i$ et 0 sinon.

On a vu que :

- $Z = \mathbb{E}[X|\mathcal{B}] \Leftrightarrow Z$ $\mathcal{B}$-mesurable + $\mathbb{E}[XW] = E[ZW]$ $\forall W$ $\mathcal{B}$-mesurable borné
- $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$

$$h(Y)\mathbb{E}[X|Y] \Leftrightarrow \forall \varphi \text{ borné } \mathbb{E}[X\varphi(Y)] = \mathbb{E}[h(Y)\varphi(Y)]$$

- Si Y discret :

$$\mathbb{E}[X|Y] = h(Y) \quad \text{avec} \quad h(y) = \begin{cases} \mathbb{E}[X|Y=y] \text{ si } \mathbb{P}(Y=y) \neq 0 \\ 0 \quad \text{sinon} \end{cases}$$

Exemple. (cas particulier)
$X = \psi(U)$ avec $X = \psi(U)$ avec $U \sim \mathcal{U}([0,1])$.

$$\text{pour } k \in \mathbb{N}^* \quad \mathcal{B}_k = \sigma\left(\left\{U \in \left[\frac{p}{k}, \frac{p+1}{k}\right[\right\}_{p\in\mathbb{Z}}\right)$$

$\mathbb{E}[|X|] = \mathbb{E}[|\psi(U)|] = \int_0^1 |\psi(u)|du$. Si $\psi \in L^1([0,1])$ alors $X \in L^1$. Que vaut $\mathbb{E}[X|\mathcal{B}_k]$?

$$\left\{U \in \left[\frac{p}{k}, \frac{p+1}{k}\right[\right\} = \{\lfloor kU \rfloor = p\}$$

Donc $\mathcal{B}_k = \sigma(\lfloor kU \rfloor)$ (beaucoup plus agréable à traiter). $\lfloor kU \rfloor$ v.a. discrète, on peut donc appliquer l'exemple.

$$\mathbb{E}[\psi(U)|\mathcal{B}_k] = \mathbb{E}[\psi(U)|\lfloor kU \rfloor] = h(\lfloor kU \rfloor)$$

avec pour $p \in \{0, 1, ..., k-1\}$ (seules valeurs telles que $\mathbb{P}(\lfloor kU \rfloor = p) \neq 0$) :

$$h(p) = \mathbb{E}[\psi(U)|\lfloor kU \rfloor = p] = \frac{\mathbb{E}[\psi(U)\mathbb{1}_{\lfloor kU \rfloor = p}]}{\mathbb{P}(\lfloor kU \rfloor = p)} = \frac{\int_0^1 \psi(u)\mathbb{1}_{\frac{p}{k} \leqslant u < \frac{p+1}{k}}}{\mathbb{P}\left(\frac{p}{k} \leqslant U < \frac{p+1}{k}\right)}$$
$$= k\int_{\frac{p}{k}}^{\frac{p+1}{k}} \psi(u)du = \text{valeur moyenne de } \psi \text{ sur } \left[\frac{p}{k}, \frac{p+1}{k}\right[$$

Ainsi,

$$\mathbb{E}[X|\mathcal{B}_k] = h(\lfloor kU \rfloor) = \int_{\frac{\lfloor kU \rfloor}{k}}^{\frac{\lfloor kU \rfloor+1}{k}} \psi(u)\frac{du}{\frac{1}{k}} = \text{valeur moyenne de } \psi \text{ sur } \left[\frac{\lfloor kU \rfloor}{k}, \frac{\lfloor kU \rfloor+1}{k}\right[$$

Effet de $\mathbb{E}[.|\mathcal{B}_k]$: perte d'information. $X = \psi(U)$ on a toute la l'information sur X, mais dans $\mathbb{E}[X|\mathcal{B}_k]$ on remplace la valeur de X par sa moyenne sur un ensemble. $\mathbb{E}[X]$ v.a. constante, on a perdu toute l'information.

Reprenons cet exemple avec $\psi(x) = \frac{1}{x}$. $\mathbb{E}\left[\frac{1}{U}|\mathcal{B}_k\right]$ existe car $\frac{1}{U} \geqslant 0$ p.s. On aura encore :

$$\mathbb{E}\left[\frac{1}{U}|\mathcal{B}_k\right] = \int_{\frac{\lfloor kU \rfloor}{k}}^{\frac{\lfloor kU \rfloor+1}{k}} \frac{1}{u}\frac{du}{\frac{1}{k}} \quad \text{p.s.} \quad = \begin{cases} +\infty \text{ si } U < \frac{1}{k} \\ \int_{\frac{1}{k}}^{\frac{2}{k}} \text{ si } \frac{1}{k} \leqslant U < \frac{2}{k} \\ ... \end{cases}$$

On a donc $0 \leqslant \frac{1}{U} < +\infty$ p.s. mais $\mathbb{P}\left(\mathbb{E}\left[\frac{1}{U}|\mathcal{B}_k\right] = +\infty\right) > 0$.

Propriété 2

1. $\mathbb{E}[.|\mathcal{B}]\colon L^1 \longrightarrow L^1$ est linéaire :
$$\mathbb{E}[X+Y|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}] + \mathbb{E}[Y|\mathcal{B}] \text{ p.s.} \quad \text{et} \quad \mathbb{E}[\lambda X|\mathcal{B}] = \lambda \mathbb{E}[X|\mathcal{B}] \text{ p.s.}$$
2. Si X est L^1 et $\mathcal{B}$-mesurable $\mathbb{E}[X|\mathcal{B}] = X$ p.s.
En particulier, $\mathbb{E}[c|\mathcal{B}] = c$ et $\mathbb{E}[\varphi(Y)|Y] = \varphi(Y)$ p.s.
3. Si X est L^1 et est indépendant de $\mathcal{B}$ (signifie que $\sigma(X)$ et $\mathcal{B}$ sont indépendants) alors $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$. En particulier X et Y indépendants $\Rightarrow$ $\mathbb{E}[X|Y] = \mathbb{E}[X]$.
4. Si $\mathcal{B} \subseteq \mathcal{C}$ deux sous-tribus de $\mathcal{A}$, X L^1,
$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]|\mathcal{C}\big] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{C}]|\mathcal{B}\big] = \mathbb{E}[X|\mathcal{B}]$$
5. (inégalité de Jensen) X L^1, $\varphi(X)$ L^1, φ convexe, alors :
$$\varphi(\mathbb{E}[X|\mathcal{B}]) \leqslant \mathbb{E}[\varphi(X)|\mathcal{B}]$$
(cas particulier utile : $\big|\mathbb{E}[X|\mathcal{B}]\big| \leqslant \mathbb{E}\big[|X|\big|\mathcal{B}\big]$)
6. Soit X, Y L^1, alors :
$$X \leqslant Y \text{ p.s.} \quad \Rightarrow \quad \mathbb{E}[X|\mathcal{B}] \leqslant \mathbb{E}[Y|\mathcal{B}] \text{ p.s.}$$
7. (crucial) X L^1, XY L^1 et Y $\mathcal{B}$-mesurable, alors :
$$\mathbb{E}[XY|\mathcal{B}] = Y\mathbb{E}[X|\mathcal{B}]$$

Remarques.

- (2) $\Rightarrow$ $\mathbb{E}[X|\mathcal{A}] = X$
- (3) $\Rightarrow$ $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}[X]$

$\mathbb{E}[X|\mathcal{B}]$ est une interpolation entre ces deux extrêmes : fait une “moyenne partielle”.

Remarque. Certaines de ces propriétés sont triviales pour des v.a. L^2. Sur L^2 on a vu que $\mathbb{E}[.|\mathcal{B}] = \Pi_{\mathcal{L}^2(\mathcal{B})}$ qui est un opérateur de projection donc on obtient directement (1) et (2).
Si $\mathcal{B} \subseteq \mathcal{C}$ alors $\mathcal{L}^2(\mathcal{B}) \subseteq \mathcal{L}^2(\mathcal{C})$ et donc :
$$\Pi_{\mathcal{L}^2(\mathcal{B})} \circ \Pi_{\mathcal{L}^2(\mathcal{C})} = \Pi_{\mathcal{L}^2(\mathcal{C})} \circ \Pi_{\mathcal{L}^2(\mathcal{B})} = \Pi_{\mathcal{L}^2(\mathcal{B})} \quad \text{donne (4)}$$

Preuve. (de la proposition)

1. Si W est $\mathcal{B}$-mesurable borné,
$$\begin{aligned}\mathbb{E}[(aX+Y)X] &= a\mathbb{E}[XW] + \mathbb{E}[YW] = a\mathbb{E}[\mathbb{E}[X|\mathcal{B}]W] + \mathbb{E}[\mathbb{E}[Y|\mathcal{B}]W] \\ &= \mathbb{E}[\quad \underbrace{(a\mathbb{E}[X|\mathcal{B}] + \mathbb{E}[Y|\mathcal{B}])}_{\mathcal{B}\text{-mesurable et satisfait prop carac}} \quad W]\end{aligned}$$
Donc $\mathbb{E}[aX+Y|\mathcal{B}] = a\mathbb{E}[X|\mathcal{B}] + \mathbb{E}[Y|\mathcal{B}]$ p.s.

2. Si X est $\mathcal{B}$-mesurable borné, $\mathbb{E}[XW] = \mathbb{E}[XW]$ donc X vérifie la propriété caractéristique, ainsi $\mathbb{E}[X|\mathcal{B}] = X$ p.s.

3. Soit W $\mathcal{B}$-mesurable borné

$$\mathbb{E}[XW] \underset{\text{idp}}{=} \mathbb{E}[X]\mathbb{E}[W] = \mathbb{E}[\underbrace{\mathbb{E}[X]}_{\mathcal{B}\text{-mesurable}} W]$$

alors comme $\mathbb{E}[X]$ $\mathcal{B}$-mesurable et vérifie la propriété caractéristique on obtient $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$.

4. Soit $\mathcal{B} \subseteq \mathcal{C}$.

$$\mathbb{E}[\underbrace{\mathbb{E}[X|\mathcal{B}]}_{\mathcal{B}\text{-mes donc } \mathcal{C}\text{-mes}} |\mathcal{C}] = \mathbb{E}[X|\mathcal{B}] \text{ p.s. d'après (2)}$$

$\mathbb{E}\big[\mathbb{E}[X|\mathcal{C}]|\mathcal{B}\big]$? Soit W $\mathcal{B}$-mesurable borné (donc $\mathcal{C}$-mesurable).

$$\begin{aligned}\mathbb{E}\big[\mathbb{E}[X|\mathcal{C}]\underbrace{W}_{\mathcal{C}\text{-mes}}\big] &\underset{\text{prop carac de } \mathbb{E}[.|\mathcal{C}]}{=} \mathbb{E}[XW] \\ &\underset{\text{prop carac de } \mathbb{E}[.|\mathcal{B}]}{=} \mathbb{E}[\mathbb{E}[X|\mathcal{B}]W]\end{aligned}$$

Donc $\mathbb{E}[X|\mathcal{B}]$ est $\mathcal{B}$-mesurable et vérifie la propriété caractéristique. Alors $\mathbb{E}\big[\mathbb{E}[X|\mathcal{C}]|\mathcal{B}\big] = \mathbb{E}[X|\mathcal{B}]$ p.s.

5. (Jensen) même preuve que pour $\mathbb{E}[.]$. Soit φ convexe, on a :

$$\varphi(x) = \sup\{ax + b \mid a, b \text{ t.q. } \forall y : \ ay + b \leqslant \varphi(y)\}$$

$$\begin{aligned}\varphi(\mathbb{E}[X|\mathcal{B}]) &= \sup\Big\{a\mathbb{E}[X|\mathcal{B}] + b \mid a, b \text{ t.q. } \forall y : \ ay + b \leqslant \varphi(y)\Big\} \\ &= \sup\Big\{\mathbb{E}[\underbrace{aX + b}_{\leqslant \varphi(X)}|\mathcal{B}] \mid a, b \text{ t.q. } \forall y : \ ay + b \leqslant \varphi(y)\Big\} \\ &\underset{(4)}{\leqslant} \sup\Big\{\mathbb{E}[\varphi(X)|\mathcal{B}] \mid a, b \text{ t.q. } \forall y : \ ay + b \leqslant \varphi(y)\Big\} \\ &= \mathbb{E}[\varphi(X)|\mathcal{B}] \text{ p.s.}\end{aligned}$$

6. Même preuve que dans le cas L^2. $W = \mathbb{1}_{\mathbb{E}[Y|\mathcal{B}] < \mathbb{E}[X|\mathcal{B}]}$, alors :

$$\begin{aligned}\mathbb{E}\Big[\underbrace{(X - Y)\mathbb{1}_{\mathbb{E}[Y|\mathcal{B}] < \mathbb{E}[X|\mathcal{B}]}}_{\leqslant 0 \text{ p.s.}}\Big] &= \mathbb{E}[XW] - \mathbb{E}[YW] \\ &= \mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]W\big] - \mathbb{E}\big[\mathbb{E}[Y|\mathcal{B}]W\big] \\ &= \mathbb{E}\Big[\underbrace{(\mathbb{E}[X|\mathcal{B}] - \mathbb{E}[Y|\mathcal{B}])\mathbb{1}_{\mathbb{E}[Y|\mathcal{B}] < \mathbb{E}[X|\mathcal{B}]}}_{\geqslant 0 \text{ p.s.}}\Big]\end{aligned}$$

Donc la v.a. dans la deuxième espérance est nulle p.s. donc $\mathbb{E}[X|\mathcal{B}] \leqslant \mathbb{E}[Y|\mathcal{B}]$ p.s.

7. Si $X \geqslant 0$ et $Y \geqslant 0$, Y $\mathcal{B}$-mesurable. Soit W $\mathcal{B}$-mesurable $\geqslant 0$.

$$\mathbb{E}[(XY)W] = \mathbb{E}[X \underbrace{(YW)}_{\mathcal{B}\text{-mes}\geqslant 0}] = \mathbb{E}\Big[\underbrace{(\mathbb{E}[X|\mathcal{B}]Y)}_{\mathcal{B}\text{-mes}}W\Big]$$

Donc $\mathbb{E}[X|\mathcal{B}]Y$ est $\mathcal{B}$-mesurable et satisfait la propriété caractéritique de $\mathbb{E}[XY|\mathcal{B}]$ pour les v.a. positives. Donc pour X, Y L^1, XY L^1 positifs :

$$\mathbb{E}[XY|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}]Y \ \in L^1$$

Dans le cas général :

$$\begin{aligned}
\mathbb{E}[XY|\mathcal{B}] = \mathbb{E}\Big[(X_+ - X_-)(Y_+ - Y_-)|\mathcal{B}\Big] &= \mathbb{E}[X_+Y_+|\mathcal{B}] + ... \\
&= \mathbb{E}[X_+|\mathcal{B}]Y_+ + ... \\
&= \mathbb{E}[(X_+ - X_-)|\mathcal{B}](Y_+ - Y_-) = \mathbb{E}[X|\mathcal{B}]Y
\end{aligned}$$

□

Exemple. (cas à densité)
Soit $(X, Y) \in \mathbb{R}^2$, $d\mu_{(X,Y)}(x, y) = \rho(x, y)dxdy$. Soit $\psi \colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ mesurable. $\mathbb{E}[|\psi(X, Y)|] < +\infty$.
Que vaut $\mathbb{E}[\psi(X, Y)|Y]$?
h caractérisé par :

$$\forall \varphi \text{ mesurable bornée} \quad \mathbb{E}[\psi(X, Y)\varphi(Y)] = \mathbb{E}[h(Y)\psi(Y)]$$

Or

$$\begin{aligned}
\mathbb{E}[\psi(X, Y)\varphi(Y)] &= \int_{\mathbb{R}^2} \psi(x, y)\varphi(y)\rho(x, y)dxdy \\
&= \int_{y\in\mathbb{R}} \varphi(y)\frac{\int_{x\in\mathbb{R}} \psi(x, y)\rho(x, y)dx}{\int_{x\in\mathbb{R}} \rho(x, y)dx}\mathbb{1}_{\int_{x\in\mathbb{R}} \rho(x,y)dx\neq 0}\left(\int_{\mathbb{R}} \rho(x, y)dx\right) dy \quad (*)
\end{aligned}$$

D'un autre côté on a :

$$\mathbb{E}[h(Y)\varphi(Y)] = \int h(y)\varphi(y)d\mu_Y(y) = \int h(y)\varphi(y)\left(\int_{\mathbb{R}} \rho(x, y)dx\right) dy$$

Ainsi on a comme expression de h :

$$h(y) = \frac{\int_{x\in\mathbb{R}} \psi(x, y)\rho(x, y)dx}{\int_{x\in\mathbb{R}} \rho(x, y)dx}\mathbb{1}_{\int_{x\in\mathbb{R}} \rho(x,y)dx\neq 0}$$

$$\mathbb{E}[\psi(X, Y)\varphi(Y)] = \mathbb{E}\left[\underbrace{\frac{\int_{x\in\mathbb{R}} \psi(x, Y)\rho(x, Y)dx}{\int_{x\in\mathbb{R}} \rho(x, Y)dx}\mathbb{1}_{\int_{x\in\mathbb{R}} \rho(x,Y)dx\neq 0}}_{=\mathbb{E}[\psi(X,Y)|Y]}\varphi(Y)\right]$$

$(*)$ En rajoutant l'indicatrice on rate :

$$\int_{\mathbb{R}^2} \psi(x,y)\varphi(y)\mathbb{1}_{\int_{\mathbb{R}} \rho(x,y)dx=0}\rho(x,y)dxdy = \int_{y\in\mathbb{R}} \left(\int_{x\in\mathbb{R}} \underbrace{\psi(x,y)\varphi(y)\mathbb{1}_{\int_{\mathbb{R}} \rho(x,y)dx=0}\rho(x,y)dx}_{(**)} \right) dy$$

$(**)$: à y fixé, soit $\int_{\mathbb{R}} \rho(x,y)dx > 0 \Rightarrow = 0$ soit $\rho(x,y) = 0$ dx-pp donc $= 0$.

Exemple. Soit $\mathcal{B}$ une tribu, X v.a. indépendante de $\mathcal{B}$, Y $\mathcal{B}$-mesurable et $\mathbb{E}[|\psi(X,Y)|] < +\infty$. Que vaut $\mathbb{E}[\psi(X,Y)|\mathcal{B}]$?
Soit W $\mathcal{B}$-mesurable borné. X est indépendant de (Y,W).

$$\begin{aligned}
\mathbb{E}[\psi(X,Y)W] &\underset{\text{transfert}}{=} \int \psi(x,y)w \underbrace{d\mu_{(X,Y,W)}(x,y,w)}_{=d\mu_X(x)d\mu_{(Y,W)}(y,w)} \\
&= \int \left(\int \psi(x,y)d\mu_X(x) \right) w d\mu_{(Y,W)}(y,w) \\
&\underset{\text{transfert}}{=} \mathbb{E}\left[\left(\int \psi(x,Y)d\mu_X(x) \right) W \right]
\end{aligned}$$

Ainsi $\int \psi(x,Y)d\mu_X(x)$ est $\mathcal{B}$-mesurable (fonction de Y donc est Y-mesurable) et satisfait la propriété caractéristique. Alors :

$$\mathbb{E}[\psi(X,Y)|\mathcal{B}] = \int \psi(x,Y)d\mu_X(x) = h(Y) \quad \text{où } h(y) = \mathbb{E}[\psi(X,y)]$$

On a les mêmes résultats de convergence que pour l'espérance.

Théorème 1

- (Fatou) Si $X_n \geqslant 0$ p.s. alors $\mathbb{E}[\underline{\lim} X_n|\mathcal{B}] \leqslant \underline{\lim}\mathbb{E}[X_n|\mathcal{B}]$
- (TCM) Si $X_n \geqslant 0$ p.s. et $X_n \uparrow X$ p.s. alors $\lim \uparrow \mathbb{E}[X_n|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}]$ p.s.
- (TCD) Si $|X_n| \leqslant Z$ p.s., $Z \in L^1$ et $X_n \xrightarrow[n\to+\infty]{} X$ p.s. alors $\mathbb{E}[X_n|\mathcal{B}] \xrightarrow[n\to+\infty]{} \mathbb{E}[X|\mathcal{B}]$ p.s et L^1.

Preuve.

- (TCM) Soit W $\mathcal{B}$-mesurable $\geqslant 0$.

$$\begin{aligned}
\mathbb{E}[XW] &= \mathbb{E}[\lim \uparrow X_n] \\
&\underset{\text{TCM}}{=} \lim \uparrow \mathbb{E}[X_nW] \\
&\underset{\text{prop carac}}{=} \lim \uparrow \mathbb{E}[\mathbb{E}[X_n|\mathcal{B}]W] \\
&\underset{\text{TCM}}{=} \mathbb{E}\Big[\underbrace{\lim \uparrow \mathbb{E}[X_n|\mathcal{B}]}_{(*)}W\Big]
\end{aligned}$$

$(*)$ $\mathcal{B}$-mesurable et satisfait la propriété caractéristique de $\mathbb{E}[X|\mathcal{B}]$. D'où le résultat.

- (Fatou)

$$\mathbb{E}[\underline{\lim} X_n|\mathcal{B}] = \mathbb{E}\Big[\lim_{n\to+\infty}\uparrow \inf_{k\geqslant n} X_k|\mathcal{B}\Big] \underset{\text{TCM conditionnel}}{=} \lim_{n\to+\infty}\mathbb{E}\big[\underbrace{\inf_{k\geqslant n} X_k}_{\leqslant X_k\ \forall k\geqslant n}|\mathcal{B}\big] \leqslant \underbrace{\lim_{n\to+\infty}\inf_{k\geqslant n}}_{\underline{\lim}}\mathbb{E}[X_n|\mathcal{B}]$$

- (TCD) $\quad -Z \leqslant X_n \leqslant Z$ donc $Z + X_n \geqslant 0$ p.s. Donc :

$$\mathbb{E}[\underline{\lim}(Z+X_n)|\mathcal{B}] \underset{\text{Fatou conditionnel}}{\leqslant} \underline{\lim}\mathbb{E}[Z+X_n|\mathcal{B}] = \mathbb{E}[Z|\mathcal{B}] + \underline{\lim}\mathbb{E}[X_n|\mathcal{B}]$$

Or :

$$\mathbb{E}[\underline{(}Z+X_n)|\mathcal{B}] = \mathbb{E}[(Z+\underline{\lim}X_n)|\mathcal{B}] = \mathbb{E}[Z|\mathcal{B}] + \mathbb{E}[\underline{\lim}X_n|\mathcal{B}]$$

Donc $\mathbb{E}[X|\mathcal{B}] \leqslant \underline{\lim}\mathbb{E}[X_n|\mathcal{B}]$ (i). De même en appliquant aux $-X_n$, $\mathbb{E}[-X|\mathcal{B}] \leqslant \underline{\lim}\mathbb{E}[-X_n|\mathcal{B}]$. Ainsi $\overline{\lim}\mathbb{E}[X_n|\mathcal{B}] \leqslant \mathbb{E}[X|\mathcal{B}]$ (ii). D'où $(i) + (ii)$:

$$\mathbb{E}[X_n|\mathcal{B}] \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbb{E}[X|\mathcal{B}]$$

Montrons maintenant la convergence L^1.

$$\mathbb{E}\Big[\underbrace{|\mathbb{E}[X_n|\mathcal{B}] - \mathbb{E}[X|\mathcal{B}]|}_{\longrightarrow 0 \text{ p.s.}}\Big] \underset{\text{oui par TCD}}{\longrightarrow} 0\,?$$

$X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$ donc $|X| \leqslant Z$ p.s.

$$\begin{aligned} |.| &\leqslant |\mathbb{E}[X_n|\mathcal{B}] + |\mathbb{E}[X|\mathcal{B}]| \\ &\underset{Jensen}{\leqslant} \mathbb{E}[|X_n||\mathcal{B}] + \mathbb{E}[|X||\mathcal{B}] \\ &\leqslant 2\mathbb{E}[Z|\mathcal{B}] \in L^1 \end{aligned}$$

Remarque. On a vu plein de propriétés (1 à 7) de $\mathbb{E}[.|\mathcal{B}]$ pour des v.a. L^1. Elles sont vraies pour des v.a. positives. On les obtient facilement grâce au TCM. $X \geqslant 0, Y \geqslant 0$:

$$\begin{aligned} \mathbb{E}[X+Y|\mathcal{B}] &= \mathbb{E}[\lim\uparrow \min(X_n) + \min(Y_n)|\mathcal{B}] \\ &= \lim\uparrow \mathbb{E}[\min(X_n)|\mathcal{B}] + \mathbb{E}[\min(Y_n)|\mathcal{B}] \\ &= \mathbb{E}[X|\mathcal{B}] + \mathbb{E}[Y|\mathcal{B}] \end{aligned}$$

□

4 Liens entre indépendance et orthogonalité

- On note parfois l'indépendance $\perp$. $(\mathcal{B}_i)_{i\in I} \perp \equiv$ les $\mathcal{B}_i$ sont indépendants.
  Dans L^2 on a le produit scalaire :

  $$\langle X, Y\rangle = \mathbb{E}[XY] \qquad X, Y \text{ indépendants } \Rightarrow \langle X, Y\rangle \mathbb{E}[X]\mathbb{E}[Y] \neq 0 \text{ en général}$$

  Ça a un peu plus de sens sur $A = \{X \in L^2 \mid \mathbb{E}[X] = 0\}$. Car sur A,

  $$\langle X, Y\rangle = \mathrm{Cov}(X, Y) \qquad \text{Donc } X, Y \in A \text{ indépendants } \Rightarrow \langle X, Y\rangle = 0$$

  La réciproque est fausse mais vraie si (X, Y) vecteur gaussien.

- On a vu aussi :

  $$\begin{aligned} &X \text{ indépendant de } Y \\ &\text{alors } \mathbb{E}[X|Y] = \mathbb{E}[X] \text{ p.s.} \\ &\text{donc } \mathbb{E}\Big[|X - \mathbb{E}[X]|\Big|Y\Big] = 0 \text{ p.s.} \\ &\text{donc } \Pi_{\mathcal{L}^2(\sigma(Y))}X - \mathbb{E}[X] = 0 \text{ p.s.} \\ &\text{donc } X - \mathbb{E}[X] \in \mathcal{L}^2(\sigma(Y))^\perp \end{aligned}$$

Propriété 3

Soient $\mathcal{B}, \mathcal{C}$ deux sous-tribus de $\mathcal{A}$. Alors :

$$\begin{aligned} & (i)\ \mathcal{B} \text{ et } \mathcal{C} \text{ sont indépendants} \\ \Leftrightarrow\ & (ii)\ \Big\{X \in \mathcal{L}^2(\mathcal{B}) \mid \mathbb{E}[X] = 0\Big\} \perp \Big\{X \in \mathcal{L}^2(\mathcal{C}) \mid \mathbb{E}[X] = 0\Big\} \\ \Leftrightarrow\ & (iii)\ \forall X \in \mathcal{L}^1(\mathcal{B}),\ \mathbb{E}[X|\mathcal{C}] = \mathbb{E}[X] \end{aligned}$$

Preuve.
$(i) \Rightarrow (iii)$: propriété de $\mathbb{E}[.|\mathcal{C}]$
$(ii) \Rightarrow (i)$: soit $B \in \mathcal{B}$ et $C \in \mathcal{C}$

$$X = \mathbb{1}_B - \mathbb{P}(B) \in \mathcal{L}^2(\mathcal{B}),\ \mathbb{E}[X] = 0$$

$$Y = \mathbb{1}_C - \mathbb{P}(C) \in \mathcal{L}^2(\mathcal{C}),\ \mathbb{E}[Y] = 0$$

Donc

$$\begin{aligned} 0 = \langle X, Y\rangle = \mathrm{Cov}(X, Y) = \mathrm{Cov}(\mathbb{1}_B, \mathbb{1}_C) = \mathbb{E}[\mathbb{1}_B\mathbb{1}_C] \\ \mathbb{E}[\mathbb{1}_{B\cap C}] = \mathbb{E}[\mathbb{1}_B]\mathbb{E}[\mathbb{1}_C] = \mathbb{P}(B \cap C) - \mathbb{P}(B)\mathbb{P}(C) \end{aligned}$$

Donc B indépendant de C $\forall (B, C)$. Ainsi $\mathcal{B}$ indépendant de $\mathcal{C}$.
$(iii) \Rightarrow (ii)$: $(iii) \Rightarrow \forall X \in \mathcal{L}^2(\mathcal{B})$ avec $\mathbb{E}[X] = 0$. $\mathbb{E}[X|\mathcal{C}] = 0$. Par construction de l'espérance conditionnelle sur $\mathcal{L}^2$, $\mathbb{E}[X|\mathcal{C}] = \Pi_{\mathcal{L}^2(\mathcal{C})}X$, donc $X \in \mathcal{L}^2(\mathcal{C})^\perp$. Donc :

$$\Big\{X \in \mathcal{L}^2(\mathcal{B}) \mid \mathbb{E}[X] = 0\Big\} \perp \mathcal{L}^2(\mathcal{C})$$

$\square$

Théorème 2

Si $Z = (X, Y_1, ..., Y_n)$ est un vecteur gaussien. Alors :

$$\mathbb{E}[X|(Y_1, ..., Y_n)] = \Pi^{\perp}_{\text{Vect}(1,Y_1,...,Y_n)} X$$

En particulier,

$$\exists \lambda_0, ..., \lambda_n,\ \mathbb{E}[X|(Y_1, ..., Y_n)]) = \lambda_0 + \lambda_1 Y_1 + ... + \lambda_n Y_n$$

Remarques.

- En particulier, $\mathbb{E}[X|(Y_1, ..., Y_n)]$ est une v.a. gaussienne.
- Intéressant car $\text{Vect}(1, Y_1, ..., Y_n)$ est un espace beaucoup plus petit que $\mathcal{L}^2(\sigma(Y_1, ..., Y_n))$.

Preuve.
Posons $V = \Pi^{\perp}_{\text{Vect}(1,Y_1,...,Y_n)} X = \lambda_0 + \lambda_1 Y_1 + ... + \lambda_n Y_n$ qui est $(Y_1, ..., Y_n)$-mesurable. Considérons $\tilde{Z} = (V - X, Y_1, ..., Y_n)$. Une combinaison linéaire de coefficients de $\tilde{Z}$ est de la forme :

$$\mu_0(V - X) + \sum_{i=1}^{n} \mu_i Y_i = -\mu_0 X + \lambda_0 \mu_0 + \sum (\mu_i + \lambda_i \mu_0) Y_i$$

et est donc une combinaison affine des Y_i et X, donc gaussienne car Z est un vecteur gaussien. Ainsi $\tilde{Z}$ est un vecteur gaussien.
Calculons $\text{Cov}(V, Y_i)$.

$$\mathbb{E}[V - X] = \langle \underbrace{V - X}_{\in \langle 1 \rangle^{\perp}}, 1 \rangle = 0$$

Alors :

$$\begin{aligned} \text{Cov}(X - Y, Y_i) &= \mathbb{E}[(V - X)Y_i] - \mathbb{E}[V - X]\mathbb{E}[Y_i] \\ &= \langle \underbrace{V - X}_{\in \langle Y_i \rangle^{\perp}}, Y_i \rangle = 0 \end{aligned}$$

La matrice de covariance de $\tilde{Z}$ est de la forme :

$$\begin{pmatrix} * & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & * & \\ 0 & & & \end{pmatrix} \in \text{Sym}_{(n+1)\times(n+1)}$$

Comme $\tilde{Z}$ est un vecteur gaussien on en déduit que $V - X$ est indépendant de $(Y_1, ..., Y_n)$. Donc si W est $Y_1, ..., Y_n$-mesurable borné :

$$\begin{aligned} \mathbb{E}[VW] &= \mathbb{E}[\underbrace{(V - X)}_{\text{idp de } W} W] + \mathbb{E}[XW] = \underbrace{\mathbb{E}[V - X]}_{=0} \mathbb{E}[W] + \mathbb{E}[XW] \\ &= \mathbb{E}[XW] \end{aligned}$$

V satisfait donc la propriété caractéristique de $\mathbb{E}[.|Y_1, ..., Y_n]$. $\square$

Exemple. Prenons :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & 5 & 3 \\ -1 & 3 & 4 \end{pmatrix} \right)$$

Que vaut $\mathbb{E}[X|(Y,Z)]$?

Posons :

$$\begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & 5 & 3 \\ -1 & 3 & 4 \end{pmatrix} \right)$$

$$\begin{aligned} \mathbb{E}[X|(Y,Z)] &= \mathbb{E}[X_0 + 1|(Y_0, Z_0)] \quad \text{car } \sigma(Y,Z) = \sigma(Y_0, Z_0) \\ &= 1 + \mathbb{E}[X|(Y_0, Z_0)] \end{aligned}$$

Par le théorème, $\mathbb{E}[X_0|(Y_0, Z_0)] = \lambda_0 + \lambda_1 Y_0 + \lambda_2 Z_0$. Pour trouver tous les coefficients on regarde $\mathbb{E}[X_0]$, $\mathbb{E}[X_0 Y_0]$ et $\mathbb{E}[X_0 Z_0]$.

- On a $0 = \mathbb{E}[X_0] = \mathbb{E}[\mathbb{E}[X_0|(Y_0, Z_0)]] = \mathbb{E}[\lambda_0 + \lambda_1 Y_0 + \lambda_2 Z_0] = \lambda_0$. Ainsi $\lambda_0 = 0$.
- $\mathbb{E}[X_0 Y_0] = \mathrm{Cov}(X_0, Y_0) = 0$. Or $\mathbb{E}[X_0 Y_0] = \mathbb{E}[(\lambda_1 Y_0 + \lambda_2 Z_0) Y_0] = \lambda \mathrm{Var} Y_0 + \lambda_2 \mathrm{Cov}(Z_0, Y_0) = 5\lambda_1 + 3\lambda_2$. Donc $5\lambda_1 + 3\lambda_2 = 0$.
- $\mathbb{E}[X_0 Z_0] = \mathrm{Cov}(X_0, Z_0) = -1$. Or $\mathbb{E}[X_0 Z_0] = \mathbb{E}[(\lambda_1 Y_0 + \lambda_2 Z_0) Z_0] = ... = 3\lambda_1 + 4\lambda_2$. Donc $3\lambda_1 + 4\lambda_2 = -1$.

Finalement on trouve $\lambda_1 = \frac{3}{1}1$ et $\lambda_2 = -\frac{5}{11}$. Ainsi :

$$\mathbb{E}[X|(Y,Z)] = 1 + \frac{3}{11} Y - \frac{5}{11}(Z-2) = \frac{21}{11} + \frac{3}{11} Y - \frac{5}{11} Z$$

5 Lois conditionnelles

Plus générale que l'espérance conditionnelle mais moins maniable.

Définition 2 (noyau de transition)

On dit que $\nu \colon \left\{ \begin{array}{ccc} \mathbb{R} \times \mathcal{B}(\mathbb{R}) & \longrightarrow & [0,1] \\ y, A & \longmapsto & \nu_y(A) \end{array} \right.$ est un noyau de transition si :

- $\forall A$, $y \mapsto v_y(A)$ est mesurable
- $\forall y$, $A \mapsto v_y(A)$ est une mesure de probabilité

$(\nu_y)_{y \in \mathbb{R}}$ famille de probabilités.

Définition 3

Si X, Y sont deux v.a. réelles, on dit que la loi de X conditionnellement à Y est (le noyaux de transition) ν si $\forall \phi$ mesurable positive :

$$\mathbb{E}[\phi(X)|Y] = \int \phi(x) d\nu_Y(x) \text{ p.s.}$$

On note $\mathcal{L}(X|Y) = \nu_Y$.

Remarques.

- On admet que (X, Y) étant donné, $\mathcal{L}(X, Y)$ existe.
- Généralise $\mathbb{E}[X|Y]$: si on sait que $\mathbb{E}[(X_+)|Y]$ et $\mathbb{E}[(X_-)|Y]$ sont finies p.s. on aura $\mathbb{E}[X|Y] = \int x d\nu_Y(x)$

Exemple. Si X indépendant de Y :

$$\mathbb{E}[\phi(X)|Y] = \mathbb{E}[\phi(X)] = \int \phi(x) d\mu_X(x)$$

Donc $\mathcal{L}(X|Y)$ existe et $\mathcal{L}(X|Y) = \mu_X$ $(\forall y,\ \nu_y = \mu_X)$.

Exemple. Si $X = h(Y)$.

$$\mathbb{E}[\phi(X)|Y] = \mathbb{E}[\phi(h(Y))|Y] = \phi(h(Y)) = \int \phi(x) d\delta_{h(Y)}(x)$$

$\mathcal{L}(X|Y) = \delta_{h(Y)}$

Exemple. $(X, Y_1, ..., Y_n)$ vecteur gaussien. Que vaut $\mathcal{L}(X|Y_1, ..., Y_n)$?

$$\mathbb{E}[\phi(X)|Y_1, ..., Y_n] = \int \phi(x) d\nu_{(Y_1,...,Y_n)}(x) \ ?$$

$$\begin{aligned}\mathbb{E}[\phi(X)|Y_1, ..., Y_n] &= \mathbb{E}[\phi(V + \tilde{V})|(Y_1, ..., y_n)] \\ &= \int \phi(V + x) d\mu_{\mathcal{N}(0,\sigma^2)}(x) \quad y = V + x \\ &= \int \phi(y) d\mu_{\mathcal{N}(V,\sigma^2)}(x)\end{aligned}$$

Avec $V = \mathbb{E}[X|Y_1, ..., Y_n]$ $Y_1, ..., Y_n$-mesurable et $\tilde{V} = X - \mathbb{E}[X|Y_1, ..., Y_n]$.
$\tilde{V} \sim \mathcal{N}(0, \sigma^2)$ gaussienne indépendant de $Y_1, ..., Y_n$. On a vu :

$$\mathbb{E}[\underbrace{f(X, Y)}_{\text{idp de } \mathcal{B}, \mathcal{B}\text{-mes}} |\mathcal{B}] = \int f(x, Y) d\mu_X(x)$$

Donc

$$\mathcal{L}(X, (Y_1, ..., Y_n)) = \mathcal{N}(\mathbb{E}[X|Y_1, ..., Y_n], \underbrace{\sigma^2}_{\text{se calcule}})$$