Esp®rances conditionnelles

2 Espérances conditionnelles

Cette notion sert à modéliser la réponse à la question suivante : si X est une *v.a.r.* liée à une certaine expérience, que sait-on d'elle si l'on n'a pas toute l'information (donnée par la tribu $\mathscr{A}$ des événements, mais seulement une information partielle (donnée par une sous-tribu $\mathscr{B}$?

2.1 Définition

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace de probabilité, et soit $\mathscr{B}$ une sous-tribu de $\mathscr{A}$; on veut définir, à partir d'une variable aléatoire réelle $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$, une autre *v.a.r.*, qui va "oublier" tout se qui se passe en dehors de $\mathscr{B}$.

Notons $\mathbb{P}_{|\mathscr{B}}$ la restriction à la sous-tribu $\mathscr{B}$ de la probabilité $\mathbb{P}$, et considérons l'espace $L^1(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}})$.

Pour tout $B \in \mathscr{B}$, on pose :

$$\nu_X(B) = \mathbb{E}\,(X \mathbb{1}_B) = \int_B X \, d\mathbb{P} \,. \tag{2.1}$$

On obtient une mesure (réelle) sur $(\Omega, \mathscr{B})$, qui est visiblement absolument continue par rapport à $\mathbb{P}_{|\mathscr{B}}$:

$$\mathbb{P}_{|\mathscr{B}}(B) = \mathbb{P}\,(B) = 0 \quad \Longrightarrow \quad \nu_X(B) = 0 \,.$$

Le Théorème de Radon-Nikodym assure donc l'existence d'une variable aléatoire $\mathscr{B}$-mesurable *unique* $Y_X \in L^1(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}})$ telle que $\nu_X = Y_X . \mathbb{P}_{|\mathscr{B}}$. On dit que Y_X est l'*espérance conditionnelle* de X par rapport à $\mathscr{B}$.

Dans toute la suite, on écrira simplement $\mathbb{P}$ au lieu de $\mathbb{P}_{|\mathscr{B}}$; on a donc :

Définition 2.1 *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace de probabilité, et soit* $\mathscr{B}$ *une sous-tribu de* $\mathscr{A}$. *Pour toute v.a.r.* $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$, *on appelle* **espérance conditionnelle** de X par rapport à $\mathscr{B}$, *ou* sachant $\mathscr{B}$, *l'unique v.a.r., notée* $\mathbb{E}^{\mathscr{B}}(X)$, *qui est* $\mathscr{B}$-mesurable *et qui vérifie :*

$$\boxed{\int_B \mathbb{E}^{\mathscr{B}}(X)\, d\mathbb{P} = \int_B X\, d\mathbb{P}}\,, \quad \boxed{\forall B \in \mathscr{B}}\,.$$

Ecrit autrement : $\boxed{\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(X)\mathbb{1}_B\right] = \mathbb{E}\left(X\mathbb{1}_B\right)}$.

On note aussi l'espérance conditionnelle $\mathbb{E}\left(X \mid \mathscr{B}\right)$.

Il résulte de la définition, en prenant $B = \Omega$ que l'on a :

Proposition 2.2 *Pour toute v.a.r. intégrable* X, *on a* $\boxed{\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(X)\right] = \mathbb{E}\left(X\right)}$.

2.2 Exemples.

a) Si $\mathscr{B} = \mathscr{A}$, il est clair que $\mathbb{E}^{\mathscr{A}}(X) = X$. Cela s'interprète en disant que, puisque l'on a toute l'information, on sait tout sur X.

Plus généralement, *si* X *est* $\mathcal{B}$*-mesurable, on a* : $\boxed{\mathbb{E}^{\mathscr{B}}(X) = X}$.

b) Si $\mathscr{B} = \{\emptyset, \Omega\}$ est la tribu grossière, alors les seules *v.a.r.* qui sont $\{\emptyset, \Omega\}$-mesurables sont les constantes. De par l'égalité $\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(X)\right] = \mathbb{E}\left(X\right)$, cette constante ne peut être que $\mathbb{E}\left(X\right)$; donc $\mathbb{E}^{\{\emptyset,\Omega\}}(X) = \mathbb{E}\left(X\right)\mathbb{1}$.

Cela s'interprète en disant que, puisque l'on ne dispose d'aucune information, tout ce que l'on peut savoir sur X est sa valeur moyenne.

c) Si $\mathscr{B}$ est la *tribu engendrée par une partition* $(B_n)_{n\geqslant 1}$ (finie ou infinie) de Ω, formée de parties $\mathscr{A}$-mesurables deux-à-deux disjointes et telles que $\mathbb{P}\left(B_n\right) \neq 0$, comme les *v.a.r.* $\mathscr{B}$-mesurables sont celles qui sont *constantes sur chaque* B_n, on a, si a_n désigne cette constante :

$$\mathbb{E}^{\mathscr{B}}(X) = \sum_{n\geqslant 1} a_n\, \mathbb{1}_{B_n}\,;$$

d'où :

$$\int_{B_n} X\, d\mathbb{P} = \int_{B_n} \mathbb{E}^{\mathscr{B}}(X)\ , d\mathbb{P} = \int_{B_n} a_n\, d\mathbb{P} = a_n\mathbb{P}\left(B_n\right).$$

Donc :

$$\boxed{\mathbb{E}^{\mathscr{B}}(X) = \sum_n \left(\frac{1}{\mathbb{P}\left(B_n\right)} \int_{B_n} X\, d\mathbb{P}\right)\mathbb{1}_{B_n}}\,;$$

ainsi, sur chaque B_n, X n'est connu que par sa moyenne sur B_n. En particulier, si $\mathscr{B} = \{\emptyset, B, B^c, \Omega\}$, pour un $B \in \mathscr{A}$ tel que $0 < \mathbb{P}(B) < 1$, alors :

$$\begin{aligned} \mathbb{E}^{\mathscr{B}}(\mathbb{1}_A) &= \frac{1}{\mathbb{P}(B)} \mathbb{P}(A \cap B) \mathbb{1}_B + \frac{1}{\mathbb{P}(B^c)} \mathbb{P}(A \cap B^c) \mathbb{1}_{B^c} \\ &= \mathbb{P}(A \mid B) \mathbb{1}_B + \mathbb{P}(A \mid B^c) \mathbb{1}_{B^c}, \end{aligned}$$

où $\mathbb{P}(A \mid B)$ est la *probabilité conditionnelle* de A sachant B.

2.3 Propriétés

Proposition 2.3 *Soit $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$. Pour toute v.a.r. Z* $\mathscr{B}$-mesurable et bornée *($Z \in L^\infty(\Omega, \mathscr{B}, \mathbb{P})$), on a :*

$$\boxed{\mathbb{E}^{\mathscr{B}}(ZX) = Z\,\mathbb{E}^{\mathscr{B}}(X)} \qquad \textit{(propriété d'idéal).}$$

Corollaire 2.4 *Pour toute $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$ et toute $Z \in L^\infty(\Omega, \mathscr{B}, \mathbb{P})$, on a :*

$$\boxed{\int_\Omega Z\,\mathbb{E}^{\mathscr{B}}(X)\,d\mathbb{P} = \int_\Omega ZX\,d\mathbb{P}}\,.$$

En effet, ce n'est autre que l'égalité $\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(ZX)\right] = \mathbb{E}(ZX)$.

Preuve de la Proposition 2.3. a) Il suffit en fait de montrer l'égalité du Corollaire 2.4, car, en remplaçant dedans Z par $Z\mathbb{1}_B$, avec $B \in \mathscr{B}$, qui est encore $\mathscr{B}$-mesurable, on aura :

$$\int_B ZX\,d\mathbb{P} = \int_B Z\,\mathbb{E}^{\mathscr{B}}(X)\,d\mathbb{P}\ ;$$

mais comme $Z\,\mathbb{E}^{\mathscr{B}}(X)$ est $\mathscr{B}$-mesurable, la définition (et l'unicité) de $\mathbb{E}^{\mathscr{B}}(ZX)$ donne bien :

$$\mathbb{E}^{\mathscr{B}}(ZX) = Z\,\mathbb{E}^{\mathscr{B}}(X)\,.$$

b) Or l'égalité :

$$\int_\Omega ZX\,d\mathbb{P} = \int_\Omega Z\,\mathbb{E}^{\mathscr{B}}(X)\,d\mathbb{P}$$

est valable, par définition, pour les $Z = \mathbb{1}_B$, avec $B \in \mathscr{B}$; elle est donc aussi valable pour les *v.a.r.* étagées $\mathscr{B}$-mesurables. Par convergence dominée, puisque $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$, elle est ensuite valable pour toutes les Z $\mathscr{B}$-mesurables bornées. $\square$

Proposition 1.5 *L'espérence conditionnelle :*

$$\mathbb{E}^{\mathscr{B}}\colon L^1(\Omega, \mathscr{A}, \mathbb{P}) \to L^1(\Omega, \mathscr{B}, \mathbb{P})$$

est une application linéaire, continue*, de norme 1,* positive*, et idempotente (projecteur).*

Remarque. L'utilisation du terme “idempotent” laisse sous-entendre que l'espace $L^1(\Omega, \mathscr{B}, \mathbb{P})$ est contenu dans $L^1(\Omega, \mathscr{A}, \mathbb{P})$; ce n'est en fait pas le cas.

En effet, il y a là une petite *difficulté*. Rappelons éléments de $L^1(\Omega, \mathscr{A}, \mathbb{P})$ ne sont pas réellement des fonctions, mais des *classes de fonctions* (modulo l'égalité presque sûre), même si d'habitude on ne fait pas de distinction entre la fonction et sa classe d'équivalence.

Maintenant, si X est une variable aléatoire $\mathscr{B}$-mesurable et si X' est une variable aléatoire ($\mathscr{A}$-mesurable), qui est presque sûrement égale à X : $\mathbb{P}(X' \neq X) = 0$, alors il n'y a aucune raison que X' soit elle aussi $\mathscr{B}$-mesurable (sauf si la tribu $\mathscr{B}$ est $\mathbb{P}$-complète). En notant $\mathbb{P}_{|\mathscr{B}}$ la restriction à $\mathscr{B}$ de la probabilité $\mathbb{P}$ (définie sur $\mathscr{A}$), on doit donc distinguer :

a) la classe d'équivalence $\mathbb{P}$-*p.s.* sur $(\Omega, \mathscr{A}, \mathbb{P})$ de X

et :

b) sa classe d'équivalence $\mathbb{P}_{|\mathscr{B}}$-*p.s.* sur $(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}})$,

la première étant en général strictement plus grande que la seconde.

En d'autres termes, bien que, pour les espaces de *fonctions*, on ait :

$$\mathscr{L}^r(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}}) \subseteq \mathscr{L}^r(\Omega, \mathscr{A}, \mathbb{P}),$$

pour $1 \leqslant r \leqslant +\infty$, par contre l'espace des *classes* de fonctions $\mathscr{B}$-mesurables $L^r(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}})$ n'est **pas** contenu dans l'espace des classes de fonctions $\mathscr{A}$-mesurables $L^r(\Omega, \mathscr{A}, \mathbb{P})$.

Néanmoins, l'application qui à la $\mathbb{P}_{|\mathscr{B}}$-classe de X fait correspondre sa $\mathbb{P}$-classe est injective, et définit une **isométrie** :

$$J\colon L^r(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}}) \longrightarrow L^r(\Omega, \mathscr{A}, \mathbb{P}).$$

Par cette isométrie, on peut donc *identifier* isométriquement $L^r(\Omega, \mathscr{B}, \mathbb{P}_{|\mathscr{B}})$ à un sous-espace (fermé) de $L^r(\Omega, \mathscr{A}, \mathbb{P})$, *ce que l'on fera toujours par la suite.*

Preuve de la Proposition 2.5. a) La linéarité est facile, par unicité.

b) Montrons que $\boxed{X \geqslant 0 \ \Rightarrow \ \mathbb{E}^{\mathscr{B}}(X) \geqslant 0}$.

En effet, si $X \geqslant 0$, la mesure ν_X définie en 1.1 est positive ; donc $\mathbb{E}^{\mathscr{B}}(X) \geqslant 0$, puisque $\nu_X = \mathbb{E}^{\mathscr{B}}(X).\mathbb{P}$

c) On a alors : $\boxed{|\mathbb{E}^{\mathscr{B}}(X)| \leqslant \mathbb{E}^{\mathscr{B}}(|X|)}$, car $|X| \pm X \geqslant 0$.

Donc, puisque $\mathbb{E}\left[\mathbb{E}^{\mathscr{B}}(|X|)\right] = \mathbb{E}(|X|) = \|X\|_1$, on obtient :

$$\|\mathbb{E}^{\mathscr{B}}(X)\|_1 \leqslant \|X\|_1 .$$

d) Mais $\boxed{\mathbb{E}^{\mathscr{B}}(\mathbb{1}) = \mathbb{1}}$ et donc la norme est exactement 1.

e) Pour finir : $\mathbb{E}^{\mathscr{B}}\big[\mathbb{E}^{\mathscr{B}}(X)\big] = \mathbb{E}^{\mathscr{B}}(X)$ car $\mathbb{E}^{\mathscr{B}}(X) \in L^1(\mathscr{B}) \subseteq L^1(\mathscr{A})$ et que $\mathbb{E}^{\mathscr{B}}(X)$ est $\mathscr{B}$-mesurable. □

Proposition 2.6

1) *Pour* $1 \leqslant r \leqslant \infty$ *:*

$$X \in L^r(\Omega, \mathscr{A}, \mathbb{P}) \quad \Rightarrow \quad \mathbb{E}^{\mathscr{B}}(X) \in L^r(\Omega, \mathscr{B}, \mathbb{P}).$$

2) *Pour* $r = 2$, $\mathbb{E}^{\mathscr{B}}$ *est la* **projection orthogonale** *de* $L^2(\Omega, \mathscr{A}, \mathbb{P})$ *sur* $L^2(\Omega, \mathscr{B}, \mathbb{P})$.

On utilisera le lemme suivant.

Lemme 2.7 *Si* $X \in L^r(\Omega, \mathscr{A}, \mathbb{P})$*, avec* $1 \leqslant r < +\infty$*, alors :*

$$\boxed{|\mathbb{E}^{\mathscr{B}}(X)|^r \leqslant \mathbb{E}^{\mathscr{B}}(|X|^r)}\,.$$

Preuve de la Proposition 2.6. 1) Pour $1 \leqslant r < \infty$, cela résulte du lemme. Pour $r = \infty$:

$$|X| \leqslant \|X\|_\infty \mathbb{1} \quad \Rightarrow \quad |\mathbb{E}^{\mathscr{B}}(X)| \leqslant \mathbb{E}^{\mathscr{B}}(|X|) \leqslant \mathbb{E}^{\mathscr{B}}(\|X\|_\infty \mathbb{1}) = \|X\|_\infty\,.$$

2) Pour $r = 2$, on doit vérifier que $X - \mathbb{E}^{\mathscr{B}}(X) \perp Z$ pour toute $Z \in L^2(\Omega, \mathscr{B}, \mathbb{P})$. Mais :

$$\int_\Omega \big[X - \mathbb{E}^{\mathscr{B}}(X)\big] Z \, d\mathbb{P} = \int_\Omega XZ \, d\mathbb{P} - \int_\Omega \mathbb{E}^{\mathscr{B}}(X) Z \, d\mathbb{P} = 0$$

pour toute $Z \in L^\infty(\Omega, \mathscr{B}, \mathbb{P})$, donc pour toute $Z \in L^2(\Omega, \mathscr{B}, \mathbb{P})$ par densité. □

Remarque. Certains auteurs définissent directement, pour $X \in L^2(\Omega, \mathscr{A}, \mathbb{P})$, l'espérance conditionnelle de X sachant $\mathscr{B}$ comme la projection orthogonale de X sur $L^2(\Omega, \mathscr{B}, \mathbb{P})$, puis prolongent l'application $\mathbb{E}^{\mathscr{B}}$ à $L^1(\Omega, \mathscr{A}, \mathbb{P})$, par densité. C'est plus élémentaire, mais cela fait complètement perdre de vue le sens réel de l'espérance conditionnelle.

Preuve du Lemme 2.7. On peut supposer $r > 1$.

Pour toute $Z \in L^\infty(\Omega, \mathscr{B}, \mathbb{P})$ et tout $B \in \mathscr{B}$, on a :

$$\begin{aligned}\int_\Omega |Z \mathbb{1}_B \mathbb{E}^{\mathscr{B}}(X)| \, d\mathbb{P} &= \int_\Omega |\mathbb{E}^{\mathscr{B}}(Z \mathbb{1}_B X)| \, d\mathbb{P} \leqslant \int_\Omega \mathbb{E}^{\mathscr{B}}(|Z \mathbb{1}_B X|) \, d\mathbb{P} \\ &= \int_\Omega |Z \mathbb{1}_B X| \, d\mathbb{P} \leqslant \|Z\|_s \|\mathbb{1}_B X\|_r < +\infty\,,\end{aligned}$$

avec $\frac{1}{r} + \frac{1}{s} = 1$. Comme $L^\infty(\Omega, \mathscr{B}, \mathbb{P})$ est dense dans $L^s(\Omega, \mathscr{B}, \mathbb{P})$, cela signifie que :

$$Z \longmapsto \int_\Omega Z \mathbb{1}_B \mathbb{E}^{\mathscr{B}}(X) \, d\mathbb{P}$$

est une forme linéaire continue sur $L^s(\Omega, \mathscr{B}, \mathbb{P})$, de norme $\leqslant \|\mathbb{1}_B X\|_r$. Donc $\mathbb{1}_B \mathbb{E}^{\mathscr{B}}(X) \in L^r(\Omega, \mathscr{B}, \mathbb{P})$, et :

$$\|\mathbb{1}_B \mathbb{E}^{\mathscr{B}}(X)\|_r \leqslant \|\mathbb{1}_B X\|_r \,;$$

cela s'écrit aussi :

$$\int_B |\mathbb{E}^{\mathscr{B}}(X)|^r \, d\mathbb{P} \leqslant \int_B |X|^r \, d\mathbb{P} = \int_B \mathbb{E}^{\mathscr{B}}(|X|^r) \, d\mathbb{P} \,.$$

Mais, puisque c'est vrai pour tout $B \in \mathscr{B}$, cela entraîne :

$$|\mathbb{E}^{\mathscr{B}}(X)|^r \leqslant \mathbb{E}^{\mathscr{B}}(|X|^r) \,. \qquad \square$$

Voyons maintenant une propriété "d'emboîtement", que l'on a déjà vue dans le cas particulier de $\mathscr{B}_2 = \{\emptyset, \Omega\}$; elle s'écrit alors $\mathbb{E}\left[\mathbb{E}^{\mathscr{B}_1}(X)\right] = \mathbb{E}(X)\mathbb{1}$.

Proposition 2.8 (transitivité) *Si $\mathscr{B}_2 \subseteq \mathscr{B}_1 \subseteq \mathscr{A}$, on a :*

$$\boxed{\mathbb{E}^{\mathscr{B}_2}\left[\mathbb{E}^{\mathscr{B}_1}(X)\right] = \mathbb{E}^{\mathscr{B}_2}(X)} \,.$$

Preuve. Si $B \in \mathscr{B}_2$:

$$\int_B \mathbb{E}^{\mathscr{B}_2}\left[\mathbb{E}^{\mathscr{B}_1}(X)\right] d\mathbb{P} = \int_B \mathbb{E}^{\mathscr{B}_1}(X) \, d\mathbb{P} \,;$$

mais comme on a aussi $B \in \mathscr{B}_1$:

$$\int_B \mathbb{E}^{\mathscr{B}_1}(X) \, d\mathbb{P} = \int_B X \, d\mathbb{P} \,. \qquad \square$$

2.4 Cas d'une tribu engendrée par une variable aléatoire

Si $\mathscr{B} = \mathscr{B}_Y$ est la tribu engendrée par une *v.a.* Y, on notera :

$$\boxed{\mathbb{E}(X \mid Y)}$$

au lieu de $\mathbb{E}^{\mathscr{B}_Y}(X) = \mathbb{E}(X \mid \mathscr{B}_Y)$. On dit que c'est l'*espérance conditionnelle de X sachant Y*.

Rappelons que l'on a :

$$\mathscr{B}_Y = \{Y^{-1}(D)\,;\ D \in \mathscr{B}or(\mathbb{R}^d)\} \,.$$

Rappelons aussi que toute *v.a.* $\mathscr{B}_Y$-mesurable s'écrit comme la composée de Y et d'une fonction borélienne sur $\mathbb{R}^d$. Donc :

Proposition 2.9 *Pour toute v.a.r. $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$ et toute v.a. Y, il existe une fonction borélienne $h\colon \mathbb{R}^d \to \mathbb{R}$, $\mathbb{P}_Y$-intégrable, telle que :*

$$\mathbb{E}(X \mid Y) = h(Y) \,.$$

Par définition, pour tout borélien $D \in \mathscr{B}or(\mathbb{R}^d)$, on a donc :

$$\int_{Y^{-1}(D)} X\, d\mathbb{P} = \int_{Y^{-1}(D)} \mathbb{E}\,(X \mid Y)\, d\mathbb{P} = \int_{Y^{-1}(D)} h(Y)\, d\mathbb{P}\,,$$

ce qui s'écrit :

$$\boxed{\int_{Y\in D} X\, d\mathbb{P} = \int_D h(y)\, d\mathbb{P}_Y(y)}\,, \quad \forall D \in \mathscr{B}or(\mathbb{R}^d)\,. \tag{2.2}$$

2.5 Exemples de calcul d'espérances conditionnelles par rapport à une v.a.r.

2.5.1 Cas où Y est une v.a.r. discrète

On a :

$$\mathbb{P}_Y = \sum_{n\geqslant 1} a_n \delta_{x_n}\,, \quad \text{avec } a_n \neq 0\,,$$

la somme étant finie ou infinie.

Puisque $\mathbb{P}\,(Y \in D) = 0$ si $D \cap \{x_n\,;\ n \geqslant 1\} = \emptyset$, la formule (1.2) ci-dessus montre que l'on peut prendre $h(y) = 0$ si $y \notin \{x_n\,;\ n \geqslant 1\}$, et :

$$h(x_n) = \frac{1}{a_n} \int_{\{Y=x_n\}} X\, d\mathbb{P} = \frac{1}{\mathbb{P}\,(Y = x_n)} \int_{\{Y=x_n\}} X\, d\mathbb{P}\,.$$

On obtient :

$$\mathbb{E}\,(X \mid Y) = \sum_{n\geqslant 1} \left(\frac{1}{\mathbb{P}\,(Y = x_n)} \int_{\{Y=x_n\}} X\, d\mathbb{P} \right) \mathbb{1}_{\{Y=x_n\}}\,,$$

ce que l'on savait déjà, puisque $\mathscr{B}_Y$ est engendrée par la partition de Ω formée par les $\{Y = x_n\}$, $n \geqslant 1$.

2.5.2 Cas de variables à densité

Supposons que le **couple** (X, Y) possède une densité $f_{(X,Y)}$ sur $\mathbb{R}^{d+1}$.

La loi $\mathbb{P}_Y$ de Y a alors une densité (*densité marginale*), donnée par :

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dx\,.$$

Comme $f_Y \in L^1(\mathbb{R}^d)$, f_Y est finie λ_d-presque partout ; par conséquent, Y est presque sûrement à valeurs dans :

$$Q = \{y \in \mathbb{R}^d\,;\ 0 < f_Y(y) < +\infty\}\,.$$

Pour $y \in Q$, on peut poser :

$$\boxed{f_X(x \mid y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}\,\cdot}$$

La fonction :

$$f_X(\,.\mid y)\colon x \longmapsto f_X(x\mid y)$$

est une *densité de probabilité* sur $\mathbb{R}$: on a $f_X(x\mid y) \geqslant 0$, et :

$$\int_{\mathbb{R}} f_X(x\mid y)\,dx = \frac{1}{f_Y(y)} \int_{\mathbb{R}} f_{(X,Y)}(x,y)\,dx = 1\,.$$

Définition 2.10 *On dit que $f_X(\,.\mid y)$ est la* **densité conditionnelle** *de X sachant $Y = y$.*

Il faut faire attention que $Y = y$ n'est qu'une notation ; en effet, $\mathbb{P}\,(Y = y) = 0$ pour tout $y \in \mathbb{R}^d$ puisque Y possède une densité.

On note parfois aussi :

$$f_X(x\mid y) = f_X(x\mid Y = y)$$

et :

$$f_X(x\mid y)\,dx = \mathbb{P}\,(x\mid Y = y)\,.$$

Proposition 2.11 *Si le vecteur aléatoire (X,Y) a une densité sur $\mathbb{R}^{d+1}$, alors pour toute $g \in L^1(\mathbb{R},\mathbb{P}_X)$, on a :*

$$\boxed{\mathbb{E}\left(g(X)\mid Y\right) = h(Y)}\,,$$

avec, pour tout pour $y \in Q$:

$$\boxed{h(y) = \int_{\mathbb{R}} g(x) f_X(x\mid y)\,dx = \frac{1}{f_Y(y)} \int_{\mathbb{R}} g(x) f_{(X,Y)}(x,y)\,dx}\,.$$

Preuve. Nous avons vu que :

$$\mathbb{E}\left(g(X)\mid Y\right) = h(Y)\,,$$

avec :

$$\int_D h(y)\,d\mathbb{P}_Y(y) = \int_{Y\in D} g(X)\,d\mathbb{P}\,, \quad \forall D \in \mathscr{B}or(\mathbb{R}^d)\,.$$

Comme $\mathbb{P}_Y(Q) = 1$, on peut poser $f_X(x\mid y) = 0$ pour $y \in \mathbb{R}^d \setminus Q$. On a alors, par le Théorème de Fubini :

$$\begin{aligned}
\int_D \left[\int_R g(x) f_X(x \mid y)\, dx \right] d\mathbb{P}_Y(y) &= \int_D \left[\int_{\mathbb{R}} g(x) f_X(x \mid y)\, dx \right] f_Y(y)\, dy \\
&= \iint_{\mathbb{R}\times D} g(x) f_X(x \mid y) f_Y(y)\, dx dy \\
&= \iint_{\mathbb{R}\times D} g(x) f_{(X,Y)}(x,y)\, dx dy \\
&= \iint_{\mathbb{R}\times D} g(x)\, d\mathbb{P}_{(X,Y)}(x,y) \\
&= \iint_{\mathbb{R}^{d+1}} g(x) \mathbb{1}_D(y)\, d\mathbb{P}_{(X,Y)}(x,y) \\
&= \int_\Omega g(X) \mathbb{1}_D(Y)\, d\mathbb{P} = \int_{Y\in D} g(x)\, d\mathbb{P} \quad \square
\end{aligned}$$

6. Propriétés de l'espérance conditionnelle analogues à celles de l'espérance

Soit X une variable aléatoire dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$. On note $\mathcal{B}$ une sous-tribu de $\mathcal{A}$.

– **a)** Pour tous réels a et b et toute variable aléatoire réelle X intégrable,

$$\mathbb{E}[aX + b|\mathcal{B}] = a\mathbb{E}[X|\mathcal{B}] + b;$$

et pour toutes variables aléatoires réelles X_1, X_2 intégrables

$$\mathbb{E}[X_1 + X_2|\mathcal{B}] = \mathbb{E}[X_1|\mathcal{B}] + \mathbb{E}[X_2|\mathcal{B}]$$

– **b)** Si $X_1 \leq X_2$ p.s. alors $\mathbb{E}[X_1|\mathcal{B}] \leq \mathbb{E}[X_2|\mathcal{B}]$.

Démonstration. Le point **a**) est une conséquence de l'unicité de l'espérance conditionnelle.

Pour le dernier point, on commence par montrer que si $X \geq 0$ alors $\mathbb{E}[X|\mathcal{B}] \geq 0$. C'est un corollaire de la preuve précédente : si $Y = \mathbb{E}[X|\mathcal{B}]$ vérifie $P(Y < 0) > 0$, alors $\mathbb{E}[Y 1_{Y<0}] < 0$; mais $1_{Y<0}$ est $\mathcal{B}$ mesurable, donc cette quantité est aussi égale à $\mathbb{E}[X 1_{Y<0}]$ qui est positive, d'où la contradiction. □

– **c)** Si X et X_n sont des variables aléatoires réelles dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$ alors

$$X_n \uparrow X \quad \Rightarrow \quad \mathbb{E}[X_n|\mathcal{B}] \uparrow \mathbb{E}[X|\mathcal{B}].$$

– **d)** Si X_n sont des variables aléatoires positives, alors

$$\mathbb{E}[\liminf X_n|\mathcal{B}] \leq \liminf \mathbb{E}[X_n|\mathcal{B}]$$

– **e)** Si $X_n \to X$ p.s. avec pour tout n, $|X_n| \leq Z \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, alors

$$\lim \mathbb{E}[X_n|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}].$$

– **f)** Soit f une fonction continue et convexe et X une variable aléatoire réelle telle que X et $f(X)$ sont intégrables, alors

$$f(\mathbb{E}[X|\mathcal{B}]) \leq \mathbb{E}[f(X)|\mathcal{B}];$$

– **g)** En particulier $|\mathbb{E}[X|\mathcal{B}]| \leq \mathbb{E}[|X||\mathcal{B}]$, et par conséquent $\mathbb{E}\left[|\mathbb{E}[X|\mathcal{B}]|\right] \leq \mathbb{E}[|X|]$.

Démonstration. Pour le point **c**) : on pose $Y = \lim \mathbb{E}[X_n|\mathcal{B}] = \limsup \mathbb{E}[X_n|\mathcal{B}]$ (d'après la croissance) qui est $\mathcal{B}$-mesurable. On a pour tout $B \in \mathcal{B}$

$$\begin{aligned}
\mathbb{E}[Y 1_B] &= \lim \mathbb{E}[\mathbb{E}[X_n|\mathcal{B}] 1_B] \quad \text{par convergence monotone} \\
&= \lim \mathbb{E}[X_n 1_B] \\
&= \mathbb{E}[X 1_B] \quad \text{par convergence monotone}
\end{aligned}$$

Pour le point **d**) : On a d'après le résultat précédent

$$\begin{aligned}
\mathbb{E}[\liminf X_n|\mathcal{B}] &= \lim_{n\to+\infty} \mathbb{E}[\inf_{k\geq n} X_k|\mathcal{B}] \\
&\leq \lim_{n\to+\infty} \inf_{k\geq n} \mathbb{E}[X_k|\mathcal{B}] = \liminf \mathbb{E}[X_n|\mathcal{B}].
\end{aligned}$$

Pour les derniers points (convergence dominée conditionnelle et inégalité de Jensen conditionnelle) il suffit de reprendre les démonstrations faites dans le cas de l'espérance classique. □

7. Propriétés spécifiques à l'espérance conditionnelle

- **a)** Si X est intégrable alors $\mathbb{E}[X|\mathcal{B}]$ l'est aussi et $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[X]$.
- **b)** Si X est une variable aléatoire réelle $\mathcal{B}$-mesurable alors

$$\mathbb{E}[X|\mathcal{B}] = X \text{ p.s.}$$

  en particulier $\mathbb{E}[1|\mathcal{B}] = 1$.
  Donc si ψ est une fonction mesurable telle que $\psi(Y)$ est intégrable, $\mathbb{E}[\psi(Y)|Y] = \psi(Y)$.
- **c)** Soient X et Z deux variables aléatoires réelles intégrables telles que XZ soit aussi intégrable. Supposons Z $\mathcal{B}$-mesurable alors

$$\mathbb{E}[XZ|\mathcal{B}] = Z\mathbb{E}[X|\mathcal{B}] \text{ p.s.};$$

  En particulier si ψ est une fonction mesurable telle que $\psi(Y)$ et $X\psi(Y)$ soient intégrables, $\mathbb{E}[X\psi(Y)|Y] = \psi(Y)\mathbb{E}[X|Y]$
- **d)** Si $\mathcal{B}_1$ et $\mathcal{B}_2$ sont deux sous-tribus de $\mathcal{A}$ telles que $\mathcal{B}_1 \subset \mathcal{B}_2$ alors

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{B}_1]|\mathcal{B}_2\right] = \mathbb{E}[X|\mathcal{B}_1] \quad \text{et} \quad \mathbb{E}\left[\mathbb{E}[X|\mathcal{B}_2]|\mathcal{B}_1\right] = \mathbb{E}[X|\mathcal{B}_1].$$

- **e)** Soit $\mathcal{B}$ une sous-tribus de $\mathcal{A}$. Si X est une variable aléatoire dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$, telle que $\sigma(X)$ et $\mathcal{B}$ sont indépendantes, alors

$$\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X];$$

  en particulier si X et Y sont deux variables aléatoires réelles indépendantes telles que $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ alors

$$\mathbb{E}[X|Y] = \mathbb{E}[X]$$

  la réciproque de ce dernier point étant fausse.

Démonstration. Les points **a)** et **b)** se déduisent de la définition.
Lorsque Z est bornée, le point **c)** se déduit de la définition de l'espérance conditionnelle et de son unicité. Puis on applique la machine standard, qui montre que la propriété est vérifiée si Z est une indicatrice, ensuite une fonction en escalier positive, puis une fonction positive $\mathcal{B}$-mesurable et enfin une fonction, Z, $\mathcal{B}$-mesurable telle que XZ soit intégrable.

Le point suivant est laissé en exercice. Enfin pour le dernier point on a pour tout $B \in \mathcal{B}$:

$$\mathbb{E}[\mathbb{E}[X|\mathcal{B}]1_B] = \mathbb{E}[X1_B] = \mathbb{E}[X]\mathbb{E}[1_B] = \mathbb{E}[\mathbb{E}[X]1_B]$$

d'où par unicité de l'espérance conditionnelle on obtient le résultat. □

L'espérance conditionnelle

Département de Mathématiques et Informatique
Centre Universitaire Abdelhafid Boussouf-Mila