Espérance conditionnelle

Réalisé par Dr. A. Redjil
Département de mathématiques, UBMA, Annaba

April 15, 2021

Abstract

E-mail: a.redjil@univ-annaba.dz

Part I
Espérance conditionnelle

1 Introduction

Pour de nombreux problèmes concrets (prédiction, observation incomplète, etc.) il est important de pouvoir estimer une variable aléatoire sur laquelle on n'a qu'une information partielle. Dès lors, on comprend l'importance de la notion d'espérance conditionnelle. La définition axiomatique de cette notion est motivée par le cas discret traité dans le premier paragraphe. Le calcul explicite des espérances conditionnelles, qui est en général un problème difficile, est illustré sur plusieurs cas, dont le cas gaussien particulièrement important pour les applications.

On note $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $B \in \mathcal{A}$ un événement tel que $\mathbb{P}(B) > 0$. On peut définir une nouvelle probabilité sur $(\Omega, \mathcal{A})$, appelée probabilité conditionnelle sachant B, en posant pour tout $A \in \mathcal{A}$,

$$\mathbb{P}(A \,|B\,) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{E}[1_A 1_B]}{\mathbb{P}(B)}.$$

De même, pour toute variable aléatoire X positive ou dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$, l'espérance conditionnelle de X sachant B est définie par

$$\mathbb{P}(X \,|B\,) = \frac{\mathbb{E}[X 1_B]}{\mathbb{P}(B)}.$$

Cette quantité est aussi l'espérance de X sous la probabilité $\mathbb{P}(\,|B\,)$, et elle s'interprète comme la valeur moyenne de X quand B est réalisé. En particulier, si X est discrète, on obtient (de façon assez conforme à l'intuition ?)

$$\mathbb{E}(X \,|B\,) = \sum_k x_k P(X = x_k \,|B\,).$$

Considérons maintenant une variable aléatoire Y à valeurs dans un espace E dénombrable et soit $Y \in E$

tel que $\mathbb{P}(Y = y) > 0$. Pour toute variable aléatoire $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ on peut définir, comme un cas particulier de ce qui précède,

$$\mathbb{E}(X\,|B) = \frac{\mathbb{E}[X1_B]}{\mathbb{P}(B)}.$$

Definition 1 *Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et Y une variable aléatoire discrète sur Ω. L'espérance conditionnelle de X sachant Y est la variable aléatoire réelle définie par*

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

où la fonction $\varphi : E \to \mathbb{R}$ est donnée par

$$\varphi(y) = \begin{cases} \mathbb{E}[X\,|Y = y] & \textit{si } y \textit{ est tel que } \mathbb{P}(Y = y) > 0 \\ & 0 \textit{ sinon} \end{cases}$$

En particulier si X est également discrète, $\mathbb{E}(X\,|Y = y) = \sum_k x_k P(X = x_k\,|Y = y)$.

Le choix de la valeur de φ lorsque $\mathbb{P}(Y = y) = 0$ n'a pas d'importance, puisque c'est un ensemble de probabilité nulle. En effet si on note $E' = \{y \in E / \mathbb{P}(Y = y) = 0\}$ alors

$$\mathbb{P}\left(Y \in E'\right) = \sum_{Y \in E'} \mathbb{P}(Y = y) = 0.$$

Donc, si on changeait la définition de φ sur E' $celadonneraitlammevariablealatoire$E[X|$Y$] à un ensemble de mesure nulle près.

Dans le cas général, l'espérance conditionnelle sera toujours définie à un ensemble de probabilité nulle près. En comparant avec le conditionnement par rapport à un événement, on observe que l'espérance conditionnelle $\mathbb{E}[X\,|Y]$ est maintenant une variable aléatoire : c'est la variable aléatoire qui donne la valeur moyenne de X quand on connait Y: on a presque sûrement

$$\mathbb{E}[X\,|Y](\omega) = \mathbb{E}[X\,|Y = y], \text{si } Y(\omega) = y.$$

On a donc aussi les écritures suivantes, avec plus ou moins d'abus de langage... :

Lorsque Y est une variable discrète à valeurs dans E

$$\begin{aligned} \mathbb{E}[X\,|Y](\omega) &= \sum_{Y \in E} 1_{Y^{-1}(\{y\})}(\omega)\,\mathbb{E}[X\,|Y = y] = \sum_{Y \in E} 1_{Y(\omega)=y} \mathbb{E}[X\,|Y = y] \\ &= \mathbb{E}[X\,|Y = Y(\omega)]. \end{aligned}$$

Remarquons que $\mathbb{E}[X\,|Y]$ est une fonction de Y (puisqu'elle ne dépend de ! que par la valeur $Y(\omega)$) donc une variable aléatoire $\sigma(Y)$ −mesurable. Dans un sens qui sera précisé plus loin, c'est la meilleure

approximation de X par une fonction de Y.

Example 2 *Lancer d'un dé. On prend* $\Omega = \{1, 2, ..., 6\}$ *et* $\mathbb{P}(\{\omega\}) = \frac{1}{6}$ *pour tout* ω*. Soient*

$$Y(\omega) = \begin{cases} 1 \text{ si } \omega \text{ est impair} \\ 0 \text{ si } \omega \text{ est pair} \end{cases}$$

et $X(\omega) = \omega$*: Alors,*

$$\mathbb{E}[X\,|Y\,](\omega) = \begin{cases} 3 \text{ si } \omega \in \{1, 3, 5\} \\ 4 \text{ si } \omega \in \{2, 4, 6\} \end{cases}$$

ou encore

$$\mathbb{E}[X\,|Y\,] = 31_{Y=1} + 41_{Y=0}.$$

Proposition 3 *Soit* $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$*. On a*

$$\mathbb{E}[\mathbb{E}[X\,|Y\,]] \leq \mathbb{E}[|X|];$$

en particulier $\mathbb{E}[X\,|Y\,] \in L^1(\Omega, \mathcal{A}, \mathbb{P})$*.*

De plus pour toute variable aléatoire Z *bornée et* $\sigma(Y)-$*mesurable*

$$\mathbb{E}[Z\mathbb{E}[X\,|Y\,]] = \mathbb{E}[ZX].$$

En particulier $\mathbb{E}[\mathbb{E}[X\,|Y\,]] = \mathbb{E}[|X|]$*.*

Proof. D'après la définition de l'espérance conditionnelle $\mathbb{E}[X\,|Y\,]$, on a

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X\,|Y\,]] &= \sum_{Y \in E\setminus E'} \mathbb{P}(Y = y)\,|\mathbb{E}[X\,|Y = y]| \\
&= \sum_{Y \in E\setminus E'} \mathbb{P}(Y = y) \left| \frac{\mathbb{E}[X1_{Y=y}]}{\mathbb{P}(Y = y)} \right| \leq \sum_{Y \in E} \mathbb{E}\left[X\,\middle|1_{Y=y}\right] = \mathbb{E}[|X|].
\end{aligned}$$

Pour la deuxième assertion, on utilise le fait qu'on peut écrire $Z = \psi(Y)$, avec ψ une fonction bornée.

Alors,

$$\begin{aligned}
\mathbb{E}[\psi(Y)\mathbb{E}[X\,|Y\,]] &= \sum_{Y \in E\setminus E'} \mathbb{P}(Y = y)\psi(y)\mathbb{E}[X\,|Y = y] \\
&= \sum_{Y \in E\setminus E'} \mathbb{P}(Y = y)\psi(y)\frac{\mathbb{E}[X1_{Y=y}]}{\mathbb{P}(Y = y)} \\
&= \sum_{Y \in E\setminus E'} \psi(y)\mathbb{E}[X1_{Y=y}] = \sum_{Y \in E\setminus E'} \mathbb{E}[\psi(y)X1_{Y=y}] \\
&= \sum_{Y \in E\setminus E'} \mathbb{E}[\psi(Y)X1_{Y=y}] = \mathbb{E}\left[\sum_{Y \in E\setminus E'} \psi(Y)1_{Y=y}X\right] \\
&= \mathbb{E}[\psi(Y)X].
\end{aligned}$$

■

Enfin, on peut vérifier que si $Y^{'}$ est une autre variable aléatoire discrète telle que $\sigma(Y) = \sigma(Y')$, on a

$\mathbb{E}[X\,|Y\,] = \mathbb{E}[X\,|Y'\,]$ p.s.

Ceci suggère que la bonne notion de conditionnement est la notion de conditionnement par rapport à une

tribu. C'est cette notion que nous allons développer dans la suite.

1.1 Définition de l'espérance conditionnelle

2.A. Cas particulier des variables aléatoires de carré intégrable. Si Y est discrète, on peut vérifier que la définition précédente entraine que si X est de carré intégrable, alors $\mathbb{E}[X|Y]$ aussi; de plus la proposition précédente entraine alors que $(\mathbb{E}[X|Y]-X)$ est orthogonal (au sens L^2) à toute variable aléatoire Z bornée $\sigma(Y)$ mesurable.

Cela suggère une généralisation de la définition lorsque Y n'est pas forcément discrète en terme de projection orthogonale.

Avant d'énoncer le résultat, rappelons que si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$ alors $L^2(\Omega,\mathcal{B},\mathbb{P})$ s'identifie à un sous-espace fermé de $L^2(\Omega,\mathcal{A},\mathbb{P})$, à savoir l'espace des éléments de $L^2(\Omega,\mathcal{A},\mathbb{P})$ dont un représentant au moins est $\mathcal{B}-$mesurable.

Definition 4 *Si $X \in L^2(\Omega,\mathcal{A},\mathbb{P})$ et si $\mathcal{B}$ est une sous tribu de $\mathcal{A}$ alors $\mathbb{E}[X|\mathcal{B}]$ est la projection orthogonale de X sur $L^2(\Omega,\mathcal{B},\mathbb{P})$. En particulier $\mathbb{E}[X|\mathcal{B}] \in L^2(\Omega,\mathcal{B},\mathbb{P})$. Si Y est une variable aléatoire, on note $\mathbb{E}[X|Y]=\mathbb{E}[X|\sigma(Y)]$: c'est la variable aléatoire fonction de Y qui approche le mieux X au sens L^2.*

Remark 5 *si $\mathcal{B}=\{\phi,\Omega\}$, $L^2(\Omega,\mathcal{B},\mathbb{P})=\{v.a\ constantes\}$, et on a alors $\mathbb{E}[X|\mathcal{B}]=\mathbb{E}[X]$. Si on en déduit immédiatement la proposition suivante:*

Proposition 6 – *Si $X \in L^2(\Omega,\mathcal{A},\mathbb{P})$ alors*

$$\mathbb{E}\left[(X-\mathbb{E}[X|\mathcal{B}])^2\right]=\inf_{Z\in L^2(\Omega,\mathcal{B},\mathbb{P})}\mathbb{E}\left[(X-Z)^2\right].$$

– *On a pour toute variable aléatoire $Z \in L^2(\Omega,\mathcal{B},\mathbb{P})$*

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

En particulier, pour toute fonction mesurable ψ telle que $\psi(Y)$ est de carré intégrable,

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

– *L'espérance conditionnelle $\mathbb{E}[X|\mathcal{B}]$ est caractérisée parmi les variables de $L^2(\Omega,\mathcal{B},\mathbb{P})$ par $\forall B \in \mathcal{B}\ \mathbb{E}[1_B\mathbb{E}[X|\mathcal{B}]]=\mathbb{E}[1_B X]$.*

Ces propriétés suggèrent la définition dans le cas L^1.

2.B. Cas général.

Theorem 7 *Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$, et soit une variable aléatoire $X \in L^1(\Omega,\mathcal{A},\mathbb{P})$. Il existe alors une unique variable aléatoire dans $X \in L^1(\Omega,m\mathcal{A},\mathbb{P})$, $\mathcal{B}$ mesurable, notée $\mathbb{E}[X|\mathcal{B}]$, telle que*

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

De manière équivalente on a, pour toute variable aléatoire, Z, $\mathcal{B}-$mesurable et bornée

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

En particulier, si $Z=1$ on a

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

L'espérance conditionnelle par rapport une tribu est caractérisée par l'une des deux propriétés ci-dessus. L'équivalence entre les deux points est assez facile à voir. La premier point nous dit que l'on a le résultat pour toutes les indicatrices $\mathcal{B}-$mesurables. Donc par somme et passage à la limite, il est encore vrai pour les fonctions étagées puis pour les fonctions bornées et $\mathcal{B}-$mesurables. Pour la réciproque il suffit de poser $Z = 1_B$.

Ainsi, par exemple, pour démontrer un résultat du style $\mathbb{E}[X|Y] = U$, une méthode est :

– On commence par vérifier que U est L^1 et est fonction de Y

– Puis on montre

$\forall \psi$ mesurable bornée, $\mathbb{E}[X\psi(Y)] = \mathbb{E}[XU]$.

Proof. 1. Existence

On a vu plus haut le cas où $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$: dans ce cas $\mathbb{E}[X|\mathcal{B}]$ est la projection orthogonale de X sur $L^2(\Omega, \mathcal{B}, \mathbb{P})$, et on a bien l'existence.

Passons maintenant au cas général, c'est-à-dire $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. En posant classiquement $X = X^+ - X^-$, il est clair que l'on peut se ramener au cas où $X \geq 0$. Soit pour $n \in \mathbb{N}$, $X_n = X \wedge n$. D'après ce qui précède on peut prendre son espérance conditionnelle par rapport à $\mathcal{B}$, $Y_n = \mathbb{E}[X_n|\mathcal{B}]$: D'autre part, X_n tend simplement en croissant vers X. Pour les Y_n remarquons que $0 \leq Y_n \leq Y_{n+1}$ presque sûrement. Il suffit pour cela de vérifier que si $U \geq 0$ alors son espérance conditionnelle vérifie $V = \mathbb{E}[U|\mathcal{B}] \geq 0$. En effet, par l'absurde : si $\mathbb{P}(V < 0) > 0$ alors il existe $\varepsilon > 0$ tel que $\mathbb{P}(V < -\varepsilon) > 0$. Or comme $\{V < -\varepsilon\} \in \mathcal{B}$ on a

$$0 \leq \mathbb{E}[U 1_{V<-\varepsilon}] = \mathbb{E}[V 1_{V<-\varepsilon}] \leq -\varepsilon$$

ce qui est impossible.

Posons $Y = \limsup Y_n$ qui est $\mathcal{B}$ mesurable. Pour tout $B \in \mathcal{B}$ on a:

$$\begin{aligned}
\mathbb{E}[Y 1_B] &= \lim_{n\to+\infty} \mathbb{E}[Y_n 1_B] \text{ par convergence monotone} \\
&= \lim_{n\to+\infty} \mathbb{E}[X_n 1_B] \\
&= \mathbb{E}[X 1_B] \text{ par convergence monotone.}
\end{aligned}$$

2. Unicité. Soient Y et Y' deux variables aléatoires dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$, $\mathcal{B}-$mesurables, et telles que

$$\forall B \in \mathcal{B}, \mathbb{E}[Y 1_B] = \mathbb{E}[X 1_B] = \mathbb{E}[Y' 1_B].$$

Comme Y et Y' sont $\mathcal{B}-$mesurables, $B_1 = \{Y > Y'\}$ et $B_2 = \{Y' > Y\}$ le sont aussi. D'où on obtient

$$\mathbb{E}[Y 1_{B_1}] = \mathbb{E}[Y' 1_{B_1}] \text{ et } \mathbb{E}[Y 1_{B_2}] = \mathbb{E}[Y' 1_{B_2}]$$

Donc, on en déduit $(Y - Y') 1_{B_1} = 0$ p.s. et $(Y - Y') 1_{B_2} = 0$ p.s. ce qui entraîne que $Y = Y'$ p.s. ∎

2 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 $\mathbb{E}[X_1\,|\mathcal{B}] \leq \mathbb{E}[X_2\,|\mathcal{B}]$.

Proof. 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 $\mathbb{P}(Y < 0) > 0$, 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 positives dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$ alors

$$X_n \uparrow X \Longrightarrow \mathbb{E}[X_n\,|\mathcal{B}] \uparrow \mathbb{E}[X\,|\mathcal{B}].$$

–**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 $\mathbb{E}[|\mathbb{E}[X\,|\mathcal{B}]|] \leq \mathbb{E}[|X|]$.

Proof. 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 $B \in \mathcal{B}$

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

■

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.

3 Propriétés spécifiques à l'espérance conditionnelle

–**a)** Si X est intégrable alors $\mathbb{E}[X\,|\mathcal{B}]$l'est aussi et $\mathbb{E}\left[\mathbb{E}[X\,|\mathcal{B}]\right] = \mathbb{E}[X]$.

–**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

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

–**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.

Proof. 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 $B \in \mathcal{B}$:

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

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

4 Calculs d'espérance conditionnelle, loi conditionnelle

$\mathcal{B} = \{\emptyset, \Omega\}$ est la tribu triviale.

Une variable aléatoire réelle est $\mathcal{B}-$mesurable si elle est constante sur Ω. Dans ce cas, $\mathbb{E}[X\,|\mathcal{B}]$ est l'unique variable aléatoire réelle constante sur Ω et égale à $\mathbb{E}[X]$.

$\mathcal{B}$ est engendré par une partition finie ou dénombrable de Ω (cela correspond au cas où $\mathcal{B} = \sigma(Y)$ lorsque Y est une variable aléatoire discrète)

On note $\{A_i, i \in J\}$, avec J dénombrable, une partition de Ω qui engendre $\mathcal{B}$. On définit L comme étant l'ensemble des indices dans J tels que $\mathbb{P}(A_i) > 0$. Alors on a

$$\mathbb{E}[X\,|\mathcal{B}] = \sum_{i\in L} \frac{\mathbb{E}[1_{A_i}X]}{\mathbb{P}(A_i)} 1_{A_i}.$$

On retrouve bien la formule donnée au début du chapitre.

En particulier si X et Y sont toutes les deux discrètes, avec $\mathbb{P}(X = x_k, Y = y_l) = p_{kl}$ et $\mathbb{P}(Y = y_l) = q_l$,

on a

$$\mathbb{P}(X = x_k\,|Y = y_l\,) = \frac{p_{kl}}{q_l}$$

Si l est fixée, la loi de probabilité sur les (x_k) donnée par $\mu(x_k) = \frac{p_{kl}}{q_l}$ est appelée "loi conditionnelle de X sachant $Y = y_l$". On a ainsi une famille de lois de probabilité indexées par les y_l. En particulier, si est ψ une fonction bornée,

$$\mathbb{E}\left[\psi(X)\,|Y = y_l\,\right] = \sum_k \psi(x_k)\,\frac{p_{kl}}{q_l}$$

et

$$\mathbb{E}\left[\psi(X)\,|Y\,\right] = \sum_l 1_{Y=y_l}\mathbb{E}\left[\psi(X)\,|Y = y_l\,\right]$$

On a alors

$$\mathbb{E}\left[\psi(X)\right] = \sum_l \mathbb{P}(Y = y_l)\,\mathbb{E}\left[\psi(X)\,|Y = y_l\,\right]$$

Cas particulier d'une somme aléatoire de variables aléatoires.

Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes intégrables identiquement distribuées.

Soit N une variable aléatoire à valeurs dans $\mathbb{N}^*$ indépendante des X_i. On pose $S_N = \sum_{i=1}^{N} X_i$. Alors $\mathbb{E}[S_N | N] = N\mathbb{E}[X_1]$. En effet on a pour tout entier k

$$\mathbb{E}[S_N | N = k] = \frac{\mathbb{E}[S_N 1_{N=k}]}{\mathbb{P}(N = k)} = \frac{\mathbb{E}\left[\left(\sum_{i=1}^{k} X_i\right) 1_{N=k}\right]}{\mathbb{P}(N = k)} = k\mathbb{E}[X_1].$$

On en déduit entre autres $\mathbb{E}[S_N] = \mathbb{E}[N]\mathbb{E}[X_1]$. On peut calculer de manière analogue la variance de S_N.

En particulier, si les X_i sont des variables de Bernoulli de paramètre p, on dira que la loi conditionnelle de S_N sachant N est une loi binomiale de paramètres (N, p).

$\mathcal{B}$ est la tribu engendré par une variable aléatoire à densité à valeurs dans $\mathbb{R}^d$. Pour simplifier les écritures on suppose que X et Y sont des variables aléatoires à densité à valeurs dans $\mathbb{R}$, de couple à densité. On note $f_{(X,Y)}(x, y)$ la densité jointe et les densités marginales $f_{(X)}(x) = \int f_{(X,Y)}(x, y)\, dy$ et $f_{(Y)}(y) = \int f_{(X,Y)}(x, y)\, dx$. On pose alors

$$f_{X|Y}(x | y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} 1_{f_{(Y)}(y)\neq 0}.$$

On a alors:

$$\mathbb{E}[X | Y](\omega) = \varphi(Y)(\omega) = \int_{\mathbb{R}} x f_{X|Y}(x | Y(\omega))\, dx.$$

$f_{X|Y}(x | y)$ est appelée “densité conditionnelle de X sachant $Y = y$” : si $f_{(Y)}(y) \neq 0$, $x \mapsto f_{X|Y}(x | y)$ est bien une densité. Ainsi on a une famille de densités (indexées par y). Cette famille de densités permet en fait de définir la “loi conditionnelle de X sachant Y” : en effet on a bien, si ψ est une fonction borélienne bornée, l'égalité presque sûre:

$$\mathbb{E}[\psi(X) | Y] = \int_{\mathbb{R}} \psi(x) f_{X|Y}(x | Y)\, dx.$$

Réciproquement, si on connaît la densité de y et la loi conditionnelle de X sachant $Y = y$, on retrouve la densité du couple

$$f_{(X,Y)}(x, y) = f_{X|Y}(x | y) f_{(Y)}(y).$$

Remark 8 *On peut vérifier que*

$$\mathbb{P}(\{\omega, f_Y(Y(\omega)) = 0\}) = \int_{f_Y^{-1}(0)} f_{(Y)}(y)\, dy = 0...$$

Example 9 *Si* (X,Y) *est uniforme sur le disque* $D(0,1)$*, la loi de* X *sachant* Y *est uniforme sur* $\left[-\sqrt{1-Y^2},\sqrt{1-Y^2}\right]$. *En effet, si* $y\in[-1,1]$,

$$f_{X|Y}\left(x\left|y\right.\right)=\frac{1_{D(x,y)}}{f_y\left(y\right)}=\frac{1_{\left[-\sqrt{1-Y^2},1+Y^2\right]}\left(x\right)}{f_y\left(y\right)}.$$

Example 10 *Si* (X,Y) *a pour densité* $2\exp\left(-\left(x+y\right)\right)1_{0\leq x\leq y}$*, on a* $f_y\left(y\right)=2\exp\left(-y\right)\left(1-\exp\left(-y\right)\right)1_{y>0}$*, la loi de* X *sachant* $Y=y$ *a pour densité* $2\frac{\exp(-x)}{1-\exp(-y)}1_{[0,y]}\left(x\right)$.

Exemple important des vecteurs gaussiens

Soit (X_1,X_2) un vecteur gaussien de loi $N\left(M,\Gamma\right)$. Alors il existe des constantes réelles a et b telles que X_1-bX_2 est indépendant de X_2 et $\mathbb{E}\left[X_1\left|X_2\right.\right]=a+bX_2$.

En effet il suffit de choisir b tels que $Cov\left(X_1-bX_2,X_2\right)=0$, c'est-à-dire $Cov\left(X_1,X_2\right)-bVar\left(X_2\right)=0$.

Puis $\mathbb{E}\left[X_1\left|X_2\right.\right]=\mathbb{E}\left[X_1-bX_2+bX_2\right]=\mathbb{E}\left[X_1-bX_2\right]+bX_2$.

Ainsi l'espérance conditionnelle d'une composante d'un vecteur gaussien par rapport à une autre composante est une fonction linéaire de cette composante. Elle suit donc à nouveau une loi normale.