Référence1 **(Springer Undergraduate Mathematics Series) Zdzisław Brzeźniak, Tomasz Zastawniak-Basic Stochastic Processes_ A Course Through Exercises-Springer (1999)**

Remarque: Cette partie en anglais est déjà faite en classe sauf les exemples 2.2 et 2.3.

2

Conditional Expectation

Conditional expectation is a crucial tool in the study of stochastic processes. It is therefore important to develop the necessary intuition behind this notion, the definition of which may appear somewhat abstract at first. This chapter is designed to help the beginner by leading him or her step by step through several special cases, which become increasingly involved, culminating at the general definition of conditional expectation. Many varied examples and exercises are provided to aid the reader's understanding.

2.1 Conditioning on an Event

The first and simplest case to consider is that of the conditional expectation $E(\xi|B)$ of a random variable ξ given an event B.

Definition 2.1

For any integrable random variable ξ and any event $B \in \mathcal{F}$ such that $P(B) \neq 0$ the *conditional expectation* of ξ given B is defined by

$$E(\xi|B) = \frac{1}{P(B)} \int_B \xi \, dP.$$

Example 2.1

Three coins, 10p, 20p and 50p are tossed. The values of those coins that land heads up are added to work out the total amount ξ. What is the expected total amount ξ given that two coins have landed heads up?

Let B denote the event that two coins have landed heads up. We want to find $E(\xi|B)$. Clearly, B consists of three elements,

$$B = \{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}\},$$

each having the same probability $\frac{1}{8}$. (Here H stands for heads and T for tails.) The corresponding values of ξ are

$$\begin{aligned} \xi(\mathrm{HHT}) &= 10 + 20 = 30, \\ \xi(\mathrm{HTH}) &= 10 + 50 = 60, \\ \xi(\mathrm{THH}) &= 20 + 50 = 70. \end{aligned}$$

Therefore

$$E(\xi|B) = \frac{1}{P(B)} \int_B \xi \, dP = \frac{1}{\frac{3}{8}} \left(\frac{30}{8} + \frac{60}{8} + \frac{70}{8} \right) = 53\tfrac{1}{3}.$$

2.2 Conditioning on a Discrete Random Variable

The next step towards the general definition of conditional expectation involves conditioning by a discrete random variable η with possible values $y_1, y_2, \ldots$ such that $P\{\eta = y_n\} \neq 0$ for each n. Finding out the value of η amounts to finding out which of the events $\{\eta = y_n\}$ has occurred or not. Conditioning by η should therefore be the same as conditioning by the events $\{\eta = y_n\}$. Because we do not know in advance which of these events will occur, we need to consider all possibilities, involving a sequence of conditional expectations

$$E(\xi|\{\eta = y_1\}), E(\xi|\{\eta = y_2\}), \ldots .$$

A convenient way of doing this is to construct a new discrete random variable constant and equal to $E(\xi|\{\eta = y_n\})$ on each of the sets $\{\eta = y_n\}$. This leads us to the next definition.

Definition 2.2

Let ξ be an integrable random variable and let η be a discrete random variable as above. Then the *conditional expectation* of ξ given η is defined to be a random variable $E(\xi|\eta)$ such that

$$E(\xi|\eta)(\omega) = E(\xi|\{\eta = y_n\}) \quad \text{if } \eta(\omega) = y_n$$

for any $n = 1, 2, \ldots$.

Example 2.2

Three coins, 10p, 20p and 50p are tossed as in Example 2.1. What is the conditional expectation $E(\xi|\eta)$ of the total amount ξ shown by the three coins given the total amount η shown by the 10p and 20p coins only?

Clearly, η is a discrete random variable with four possible values: 0, 10, 20 and 30. We find the four corresponding conditional expectations in a similar way as in Example 2.1:

$$E(\xi|\{\eta = 0\}) = 25, \qquad E(\xi|\{\eta = 10\}) = 35,$$
$$E(\xi|\{\eta = 20\}) = 45, \qquad E(\xi|\{\eta = 30\}) = 55.$$

Therefore

$$E(\xi|\eta)(\omega) = \begin{cases} 25 & \text{if } \eta(\omega) = 0, \\ 35 & \text{if } \eta(\omega) = 10, \\ 45 & \text{if } \eta(\omega) = 20, \\ 55 & \text{if } \eta(\omega) = 30. \end{cases}$$

Example 2.3

Take $\Omega = [0,1]$ with the σ-field of Borel sets and P the Lebesgue measure on $[0,1]$. We shall find $E(\xi|\eta)$ for

$$\xi(x) = 2x^2, \qquad \eta(x) = \begin{cases} 1 & \text{if } x \in [0, \frac{1}{3}], \\ 2 & \text{if } x \in (\frac{1}{3}, \frac{2}{3}), \\ 0 & \text{if } x \in (\frac{2}{3}, 1]. \end{cases}$$

Clearly, η is discrete with three possible values $1, 2, 0$. The corresponding events are

$$\begin{aligned} \{\eta = 1\} &= [0, \tfrac{1}{3}], \\ \{\eta = 2\} &= (\tfrac{1}{3}, \tfrac{2}{3}), \\ \{\eta = 0\} &= (\tfrac{2}{3}, 1]. \end{aligned}$$

For $x \in [0, \frac{1}{3}]$

$$E(\xi|\eta)(x) = E(\xi|[0, \tfrac{1}{3}]) = \frac{1}{\frac{1}{3}} \int_0^{\frac{1}{3}} 2x^2 dx = \frac{2}{27}.$$

For $x \in (\frac{1}{3}, \frac{2}{3})$

$$E(\xi|\eta)(x) = E(\xi|(\tfrac{1}{3}, \tfrac{2}{3})) = \frac{1}{\frac{1}{3}} \int_{\frac{1}{3}}^{\frac{2}{3}} 2x^2 dx = \frac{14}{27}.$$

And for $x \in (\frac{2}{3}, 1]$

$$E(\xi|\eta)(x) = E(\xi|(\tfrac{2}{3}, 1]) = \frac{1}{\frac{1}{3}} \int_{\frac{2}{3}}^{1} 2x^2 dx = \frac{38}{27}.$$

The graph of $E(\xi|\eta)$ is shown in Figure 2.1 together with those of ξ and η.

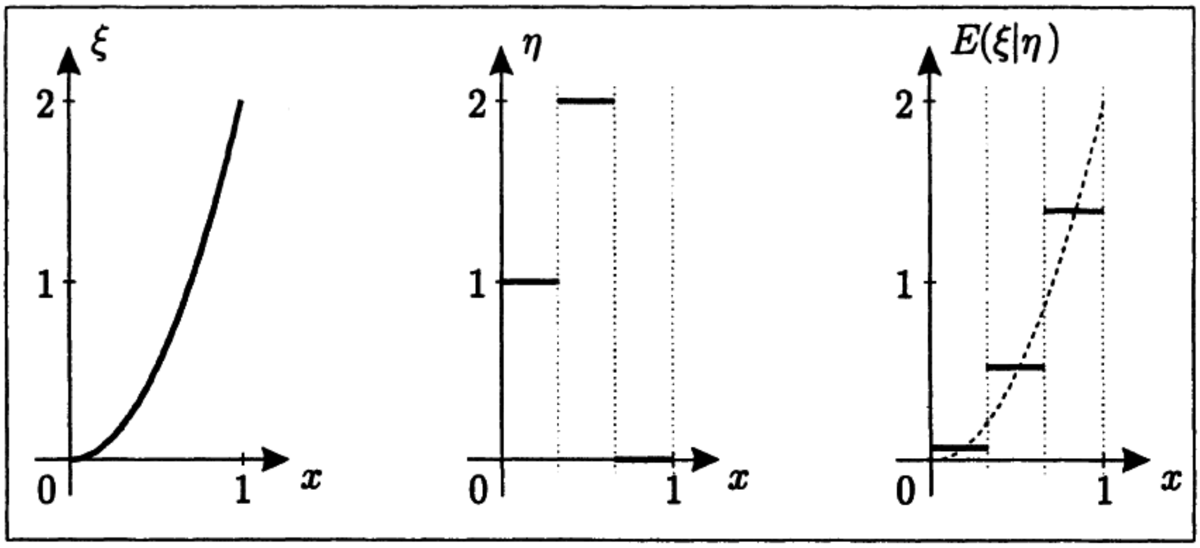

Figure 2.1. The graph of $E(\xi|\eta)$ in Example 2.3

Hint How many different values does 1_B take? What are the sets on which these values are taken?

Espérance conditionnelle par rapport à une sous-tribu et en particulier par rapport à une variable aléatoire.

Référence 2 : Nils Berglund. Martingales et calcul stochastique. DEA. Université d'Orléans, 2010, pp.91.

3.2 Espérance conditionnelle

Dans cette section, nous fixons un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et une sous-tribu $\mathcal{F}_1 \subset \mathcal{F}$. Nous avons vu que $\mathcal{F}_1$ représente une information partielle sur l'espace, obtenue par exemple en observant une variable aléatoire X_1. L'espérance conditionnelle d'une variable aléatoire X par rapport à $\mathcal{F}_1$ représente la meilleure estimation que l'on puisse faire de la valeur de X à l'aide de l'information contenue dans $\mathcal{F}_1$.

Définition 3.2.1 (Espérance conditionnelle). *Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que $\mathbb{E}(|X|) < \infty$. On appelle* espérance conditionnelle de X sachant $\mathcal{F}_1$, *et on note $\mathbb{E}(X|\mathcal{F}_1)$, toute variable aléatoire Y satisfaisant les deux conditions*

1. *$Y \subseteq \mathcal{F}_1$, c'est-à-dire Y est $\mathcal{F}_1$-mesurable;*
2. *pour tout $A \in \mathcal{F}_1$, on a*

$$\int_A X \, \mathrm{d}\mathbb{P} = \int_A Y \, \mathrm{d}\mathbb{P} \,. \tag{3.2.1}$$

Si Z est une variable aléatoire réelle sur $(\Omega, \mathcal{F}, \mathbb{P})$, nous abrégeons $\mathbb{E}(X|\sigma(Z))$ par $\mathbb{E}(X|Z)$.

En fait, toute variable aléatoire Y satisfaisant la définition est appelée une *version* de $\mathbb{E}(X|\mathcal{F}_1)$. Le résultat suivant tranche la question de l'existence et de l'unicité de l'espérance conditionnelle.

Théorème 3.2.2.

1. *L'espérance conditionnelle $\mathbb{E}(X|\mathcal{F}_1)$ existe.*
2. *L'espérance conditionnelle est unique dans le sens que si Y et Y' sont deux versions de $\mathbb{E}(X|\mathcal{F}_1)$, alors $Y = Y'$ presque sûrement.*
3. *On a $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)) = \mathbb{E}(X)$ et $\mathbb{E}(|\mathbb{E}(X|\mathcal{F}_1)|) \leqslant \mathbb{E}(|X|)$.*

DÉMONSTRATION. Commençons par prouver les assertions du point 3. La première est un cas particulier de (3.2.1) avec $A = \Omega$. Pour montrer la seconde, soit $Y = \mathbb{E}(X|\mathcal{F}_1)$ et $A = \{Y > 0\} := \{\omega \in \Omega \colon Y(\omega) > 0\}$. On a $A \in \mathcal{F}_1$ par mesurabilité de Y. Or par (3.2.1),

$$\begin{aligned} \int_A Y \, \mathrm{d}\mathbb{P} &= \int_A X \, \mathrm{d}\mathbb{P} \leqslant \int_A |X| \, \mathrm{d}\mathbb{P} \,, \\ \int_{A^c} -Y \, \mathrm{d}\mathbb{P} &= \int_{A^c} -X \, \mathrm{d}\mathbb{P} \leqslant \int_{A^c} |X| \, \mathrm{d}\mathbb{P} \,. \end{aligned} \tag{3.2.2}$$

Le résultat suit en ajoutant ces deux inégalités.

Montrons l'unicité. Si Y et Y' sont deux versions de $\mathbb{E}(X|\mathcal{F}_1)$, alors pour tout $A \in \mathcal{F}_1$

$$\int_A Y \, \mathrm{d}\mathbb{P} = \int_A Y' \, \mathrm{d}\mathbb{P} \,. \tag{3.2.3}$$

Prenons $A = \{Y - Y' \geqslant \varepsilon\}$ pour $\varepsilon > 0$. Alors

$$0 = \int_A (Y - Y') \, \mathrm{d}\mathbb{P} \geqslant \varepsilon \mathbb{P}(A) = \varepsilon \mathbb{P}\{Y - Y' \geqslant \varepsilon\} \, . \tag{3.2.4}$$

Comme c'est vrai pour tout ε, on a $Y \leqslant Y'$ presque sûrement. En interchangeant les rôles de Y et Y', on obtient l'inégalité inverse, d'où on déduit l'égalité presque sûre.

Montrons finalement l'existence. Rappelons qu'une mesure ν sur $(\Omega, \mathcal{F}_1)$ est dite *absolument continue par rapport à la mesure* μ si $\mu(A) = 0$ implique $\nu(A) = 0$ pour tout $A \in \mathcal{F}_1$. On écrit alors $\nu \ll \mu$. Le théorème de Radon–Nikodym affirme qu'il existe alors une fonction $f \subseteq \mathcal{F}_1$ telle que

$$\int_A f \, \mathrm{d}\mu = \nu(A) \qquad \forall A \in \mathcal{F}_1 \, . \tag{3.2.5}$$

La fonction f est appelée *dérivée de Radon–Nikodym* et notée $\mathrm{d}\nu/\mathrm{d}\mu$.

Supposons d'abord que $X \geqslant 0$. Posons $\mu = \mathbb{P}$ et définissons ν par

$$\nu(A) = \int_A X \, \mathrm{d}\mathbb{P} \qquad \forall A \in \mathcal{F}_1 \, , \tag{3.2.6}$$

de sorte que $\nu \ll \mu$. Nous avons $\mathrm{d}\nu/\mathrm{d}\mu \subseteq \mathcal{F}_1$ et pour tout $A \in \mathcal{F}_1$

$$\int_A X \, \mathrm{d}\mathbb{P} = \nu(A) = \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \, \mathrm{d}\mathbb{P} \, . \tag{3.2.7}$$

Ceci montre que $\mathrm{d}\nu/\mathrm{d}\mu$ satisfait (3.2.1). De plus, prenant $A = \Omega$ on voit que $\mathrm{d}\nu/\mathrm{d}\mu$ est intégrable. Par conséquent $\mathrm{d}\nu/\mathrm{d}\mu$ est une version de $\mathbb{E}(X|\mathcal{F}_1)$.

Finalement, un X général peut se décomposer $X = X^+ - X^-$ avec $X^+, X^- \geqslant 0$. Soit $Y_1 = \mathbb{E}(X^+|\mathcal{F}_1)$ et $Y_2 = \mathbb{E}(X^-|\mathcal{F}_1)$. Alors $Y_1 - Y_2$ est $\mathcal{F}_1$-mesurable et intégrable et on a pour tout $A \in \mathcal{F}_1$

$$\int_A X \, \mathrm{d}\mathbb{P} = \int_A X^+ \, \mathrm{d}\mathbb{P} - \int_A X^- \, \mathrm{d}\mathbb{P} = \int_A Y_1 \, \mathrm{d}\mathbb{P} - \int_A Y_2 \, \mathrm{d}\mathbb{P} = \int_A (Y_1 - Y_2) \, \mathrm{d}\mathbb{P} \, . \tag{3.2.8}$$

Ceci montre que $Y_1 - Y_2$ est une version de $\mathbb{E}(X|\mathcal{F}_1)$. □

Dans la suite, nous allons en général ignorer l'existence de plusieurs versions de l'espérance conditionnelle, puisque des variables aléatoires égales presque partout sont indistinguables en pratique.

Exemple 3.2.3.

1. Supposons que X soit $\mathcal{F}_1$-mesurable. Alors $\mathbb{E}(X|\mathcal{F}_1) = X$ vérifie la définition. Cela traduit le fait que $\mathcal{F}_1$ contient déjà toute l'information sur X.
2. L'autre extrême est le cas de l'indépendance. Rappelons que X est indépendante de $\mathcal{F}_1$ si pour tout $A \in \mathcal{F}_1$ et tout borélien $B \subset \mathbb{R}$,

   $$\mathbb{P}\big(\{X \in B\} \cap A\big) = \mathbb{P}\{X \in B\}\mathbb{P}(A) \, . \tag{3.2.9}$$

   Dans ce cas nous avons $\mathbb{E}(X|\mathcal{F}_1) = \mathbb{E}(X)$, c'est-à-dire qu'en l'absence de toute information, la meilleure estimation que l'on puisse faire de X est son espérance. En particulier, on notera que toute variable aléatoire est indépendante de la tribu triviale $\mathcal{F}_0$, et que par conséquent on aura toujours $\mathbb{E}(X|\mathcal{F}_0) = \mathbb{E}(X)$.

Pour vérifier la première condition de la définition, il suffit d'observer que $\mathbb{E}(X)$, étant une constante, est mesurable par rapport à la tribu triviale $\mathcal{F}_0$, donc aussi par rapport à $\mathcal{F}_1$. Pour vérifier la seconde assertion, prenons $A \in \mathcal{F}_1$. Alors $1_A \subseteq \mathcal{F}_1$ et par indépendance

$$\int_A X \, \mathrm{d}\mathbb{P} = \mathbb{E}(X 1_A) = \mathbb{E}(X)\mathbb{E}(1_A) = \mathbb{E}(X)\mathbb{P}(A) = \int_A \mathbb{E}(X) \, \mathrm{d}\mathbb{P} \, . \tag{3.2.10}$$

3. Soit $\Omega_1, \Omega_2, \dots$ une partition de Ω telle que $\mathbb{P}(\Omega_i)$ soit strictement positif pour tout i. Soit $\mathcal{F}_1 = \sigma(\Omega_1, \Omega_2, \dots)$ la tribu engendrée par les Ω_i. Alors

$$\mathbb{E}(X|\mathcal{F}_1)(\omega) = \frac{\mathbb{E}(X 1_{\Omega_i})}{\mathbb{P}(\Omega_i)} = \frac{1}{\mathbb{P}(\Omega_i)} \int_{\Omega_i} X \, \mathrm{d}\mathbb{P} \qquad \forall \omega \in \Omega_i \, . \tag{3.2.11}$$

Dans ce cas, l'information contenue dans $\mathcal{F}_1$ spécifie dans quel Ω_i on se trouve, et la meilleure estimation de X est donc sa moyenne sur Ω_i.
Pour le vérifier, observons d'abord que comme $\mathbb{E}(X|\mathcal{F}_1)$ est constante sur chaque Ω_i, elle est mesurable par rapport à $\mathcal{F}_1$. De plus,

$$\int_{\Omega_i} \mathbb{E}(X|\mathcal{F}_1) \, \mathrm{d}\mathbb{P} = \mathbb{E}(X 1_{\Omega_i}) = \int_{\Omega_i} X \, \mathrm{d}\mathbb{P} \, . \tag{3.2.12}$$

Comme $\mathcal{F}_1$ est engendrée par les Ω_i, le résultat suit par σ-additivité.

4. Considérons une chaîne de Markov $\{X_n\}_{n\in\mathbb{N}}$ sur un ensemble $\mathcal{X}$, de matrice de transition $P = (p_{i,j})$, et soit $f : \mathcal{X} \to \mathbb{R}$ une fonction mesurable. Nous prétendons que pour tout $n \in \mathbb{N}$,

$$\mathbb{E}(f(X_{n+1})|X_n) = \sum_{j\in\mathcal{X}} f(j) p_{X_n,j} \, . \tag{3.2.13}$$

En effet, cette expression étant constante sur tout ensemble où X_n est constant, elle est mesurable par rapport à $\sigma(X_n)$. De plus, en appliquant (3.2.11) avec la partition donnée par $\Omega_i = \{\omega \colon X_n = i\}$, on a pour $\omega \in \Omega_i$

$$\begin{aligned}
\mathbb{E}(f(X_{n+1})|X_n)(\omega) &= \frac{\mathbb{E}(f(X_{n+1}) 1_{\{X_n = i\}})}{\mathbb{P}\{X_n = i\}} \\
&= \sum_{j\in\mathcal{X}} f(j) \frac{\mathbb{P}\{X_{n+1} = j, X_n = i\}}{\mathbb{P}\{X_n = i\}} \\
&= \sum_{j\in\mathcal{X}} f(j) \mathbb{P}\{X_{n+1} = j | X_n = i\} = \sum_{j\in\mathcal{X}} f(j) p_{i,j} \, .
\end{aligned} \tag{3.2.14}$$

Proposition 3.2.4. *L'espérance conditionnelle a les propriétés suivantes :*

1. *Linéarité :* $\mathbb{E}(aX + Y|\mathcal{F}_1) = a\mathbb{E}(X|\mathcal{F}_1) + \mathbb{E}(Y|\mathcal{F}_1)$.
2. *Monotonie : Si* $X \leqslant Y$ *alors* $\mathbb{E}(X|\mathcal{F}_1) \leqslant \mathbb{E}(Y|\mathcal{F}_1)$.
3. *Convergence monotone : Si* $X_n \geqslant 0$ *est une suite croissante telle que* $X_n \nearrow X$ *avec* $\mathbb{E}(X) < \infty$ *alors* $\mathbb{E}(X_n|\mathcal{F}_1) \nearrow \mathbb{E}(X|\mathcal{F}_1)$.
4. *Inégalité de Jensen : Si* φ *est convexe et* $\mathbb{E}(|X|)$ *et* $\mathbb{E}(|\varphi(X)|)$ *sont finies alors*

$$\varphi(\mathbb{E}(X|\mathcal{F}_1)) \leqslant \mathbb{E}(\varphi(X)|\mathcal{F}_1) \, . \tag{3.2.15}$$

5. *Contraction dans* L^p *pour* $p \geqslant 1$ *:* $\mathbb{E}(|\mathbb{E}(X|\mathcal{F}_1)|^p) \leqslant \mathbb{E}(|X|^p)$.

Nous laissons la preuve en exercice. Le résultat suivant décrit comment les espérances conditionnelles se comportent par rapport à des sous-tribus. Il dit en résumé que c'est toujours la tribu la plus grossière qui l'emporte.

Proposition 3.2.5. *Si $\mathcal{F}_1 \subset \mathcal{F}_2$, alors*

1. $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1)$;
2. $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(X|\mathcal{F}_1)$.

DÉMONSTRATION. Pour montrer la première identité, il suffit de noter que $\mathbb{E}(X|\mathcal{F}_1) \subseteq \mathcal{F}_2$ et d'appliquer le premier point de l'exemple 3.2.3. Pour la seconde relation, on observe que pour tout $A \in \mathcal{F}_1 \subset \mathcal{F}_2$,

$$\int_A \mathbb{E}(X|\mathcal{F}_1)\,\mathrm{d}\mathbb{P} = \int_A X\,\mathrm{d}\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{F}_2)\,\mathrm{d}\mathbb{P}\,, \tag{3.2.16}$$

la première égalité suivant de $\mathbb{E}(X|\mathcal{F}_1) \subseteq \mathcal{F}_1$ et la seconde de $\mathbb{E}(X|\mathcal{F}_2) \subseteq \mathcal{F}_2$. □

Le résultat suivant montre que les variables aléatoires $\mathcal{F}_1$-mesurables se comportent comme des constantes relativement aux espérances conditionnelles par rapport à $\mathcal{F}_1$.

Théorème 3.2.6. *Si $X \subseteq \mathcal{F}_1$ et $\mathbb{E}(|Y|), \mathbb{E}(|XY|) < \infty$, alors*

$$\mathbb{E}(XY|\mathcal{F}_1) = X\mathbb{E}(Y|\mathcal{F}_1)\,. \tag{3.2.17}$$

DÉMONSTRATION. Le membre de droite étant $\mathcal{F}_1$-mesurable, la première condition de la définition est vérifiée. Pour vérifier la seconde condition, nous commençons par considérer le cas $X = 1_B$ avec $B \in \mathcal{F}_1$. Alors pour tout $A \in \mathcal{F}_1$ on a

$$\int_A 1_B\,\mathbb{E}(Y|\mathcal{F}_1)\,\mathrm{d}\mathbb{P} = \int_{A\cap B} \mathbb{E}(Y|\mathcal{F}_1)\,\mathrm{d}\mathbb{P} = \int_{A\cap B} Y\,\mathrm{d}\mathbb{P} = \int_A 1_B Y\,\mathrm{d}\mathbb{P}\,. \tag{3.2.18}$$

Ceci montre que la seconde condition est vérifiée pour des fonctions indicatrices. Par linéarité, elle est aussi vraie pour des fonctions étagées $\sum a_i 1_{B_i}$. Le théorème de convergence monotone permet d'étendre le résultat aux variables $X, Y \geqslant 0$. Enfin, pour le cas général, il suffit de décomposer X et Y en leurs parties positive et négative. □

Enfin, le résultat suivant montre que l'espérance conditionnelle peut être considérée comme une projection de $L^2(\mathcal{F}) = \{Y \subseteq \mathcal{F} \colon \mathbb{E}(Y^2) < \infty\}$ dans $L^2(\mathcal{F}_1)$.

Théorème 3.2.7. *Si $\mathbb{E}(X^2) < \infty$, alors $\mathbb{E}(X|\mathcal{F}_1)$ est la variable $Y \subseteq \mathcal{F}_1$ qui minimise $\mathbb{E}((X-Y)^2)$.*

DÉMONSTRATION. Pour tout $Z \in L^2(\mathcal{F}_1)$, le théorème précédent montre que $Z\mathbb{E}(X|\mathcal{F}_1) = \mathbb{E}(ZX|\mathcal{F}_1)$. Prenant l'espérance, on obtient

$$\mathbb{E}\big(Z\mathbb{E}(X|\mathcal{F}_1)\big) = \mathbb{E}\big(\mathbb{E}(ZX|\mathcal{F}_1)\big) = \mathbb{E}(ZX)\,, \tag{3.2.19}$$

ou encore

$$\mathbb{E}\big(Z[X - \mathbb{E}(X|\mathcal{F}_1)]\big) = 0\,. \tag{3.2.20}$$

Si $Y = \mathbb{E}(X|\mathcal{F}_1) + Z \subseteq \mathcal{F}_1$, alors

$$\mathbb{E}\big((X-Y)^2\big) = \mathbb{E}\big((X - \mathbb{E}(X|\mathcal{F}_1) - Z)^2\big) = \mathbb{E}\big((X - \mathbb{E}(X|\mathcal{F}_1))^2\big) + \mathbb{E}\big(Z^2\big)\,, \tag{3.2.21}$$

qui est minimal quand $Z = 0$. □