5 Conditional Expectations

In this chapter, our main goal is to define conditional expectations and discuss their properties. Intuitively speaking, given a $\mathcal{A}$-measurable random variable X, its conditional expectation with respect to the σ-algebra $\mathcal{B}$ can be understood as the “best” approximation of X by $\mathcal{B}$-measurable functions. When we deal with real problems, take the future prediction as an example, since at the “current moment”, we only have partial information on a “random variable describing the future”, the notion of the conditional expectation gives a mathematical way to formulate such problems.

We will start with the case of discrete random variables, give the “characteristic property” of a conditional expectation that we will use in Section 5.2 to define in an axiomatic way the conditional expectation of an integrable random variable or a non-negative random variable. Later in Section 5.4, we will also see how to compute conditional expectations and in particular, the example of the Gaussian random variables in Section 5.4.3 is of particular interest. To conclude the chapter, we will discuss the notion of conditional distribution in Section 5.5 as a preparation for the Markov chains in Chapter 7.

5.1 Case of Discrete Random Variables

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let $B \in \mathcal{A}$ be a measurable event with positive probability $\mathbb{P}(B) > 0$. We can define a new probability measure on the measurable space $(\Omega, \mathcal{A})$, called the *conditional probability* (條件機率) knowing the event B, which writes,

$$\forall A \in \mathcal{A}, \quad \mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Similarly, given a random variable X, if X is non-negative or $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, then the *conditional expectation* (條件期望值) of X knowing the event B writes,

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

We can understand the above definition as the expectation of X under the probability measure $\mathbb{P}(\cdot \,|\, B)$, which can be thought of “the average of X when B holds”.

Above we defined the conditional expectation of a random variable X knowing a measurable event. Next, we want to extend such a definition and define the conditional expectation of X knowing another random variable Y. First, consider a countable set E and a random variable Y with values in E. Let $E' = \{y \in E : \mathbb{P}(Y = y) > 0\}$. Then, for any $y \in E'$ and any random variable $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, we can define from above that,

$$\mathbb{E}[X \,|\, Y = y] = \frac{\mathbb{E}[X \mathbb{1}_{Y=y}]}{\mathbb{P}(Y = y)}.$$

Definition 5.1.1 : Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. The conditional expectation of the random variable X knowing Y is a random variable, defined as,

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

where the function $\varphi : E \longrightarrow \mathbb{R}$ satisfies,

$$\varphi(y) = \begin{cases} \mathbb{E}[X \mid Y = y], & \text{if } y \in E', \\ 0, & \text{otherwise.} \end{cases}$$

Remark 5.1.2 : In the above definition, the value of φ on $E \backslash E'$ is not relevant and does not affect the definition of $\mathbb{E}[X \mid Y]$ for the following reason,

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

Therefore, if we change the definition of φ on $E \backslash E'$, the resulting conditional expectation $\mathbb{E}[X \mid Y]$ will be the same outside a set of zero measure.

Now we note that the conditional expectation of a random variable is itself a *random variable* satisfying the following condition: knowing the value of Y, the conditional expectation gives the expectation of X under this condition. Alternatively speaking,

$$\text{if } Y(\omega) = y \in E', \text{ then } \mathbb{E}[X \mid Y](\omega) = \mathbb{E}[X \mid Y = y].$$

Moreover, we can also notice that $\mathbb{E}[X \mid Y]$ is $\sigma(Y)$-measurable being a function of Y. We will also see in a while that, among all the $\sigma(Y)$-measurable random variables, it is the "best" approximation of X.

Example 5.1.3 : Let $\Omega = \{1, \ldots, 6\}$. For all $\omega \in \Omega$, define $\mathbb{P}(\{\omega\}) = \frac{1}{6}$. Define the following random variable,

$$Y(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is odd,} \\ 0 & \text{if } \omega \text{ if even.} \end{cases}$$

Then set $X(\omega) = \omega$, and we have,

$$E[X \mid Y](\omega) = \begin{cases} 3 & \text{if } \omega \in \{1, 3, 5\}, \\ 4 & \text{if } \omega \in \{2, 4, 6\}. \end{cases}$$

Proposition 5.1.4 : *We have* $\mathbb{E}[|\,\mathbb{E}[X \mid Y]|] \leqslant \mathbb{E}[|X|]$, *which implies* $\mathbb{E}[X \mid Y] \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. *Additionally, for all bounded* $\sigma(Y)$*-measurable random variable* Z, *we have,*

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

Proof : From the definition of the conditional expectation $\mathbb{E}[X \mid Y]$, we have,

$$\mathbb{E}[|\,\mathbb{E}[X \mid Y]|] = \sum_{y \in E'} \mathbb{P}(Y = y) \frac{|\,\mathbb{E}[X \mathbb{1}_{Y=y}]|}{\mathbb{P}(Y = y)} \leqslant \sum_{y \in E'} \mathbb{E}[|X| \mathbb{1}_{Y=y}] = \mathbb{E}[|X|].$$

For the second part of the statement, we are given a $\sigma(Y)$-measurable bounded random variable Z, we can find a bounded measurable function ψ such that $Z = \psi(Y)$ from Proposition 2.2.6, giving,

$$\mathbb{E}[\psi(Y)\,\mathbb{E}[X \mid Y]] = \sum_{y \in E} \psi(y)\,\mathbb{E}[X \mathbb{1}_{Y=y}] = \sum_{y \in E} \mathbb{E}[\psi(Y) X \mathbb{1}_{Y=y}] = \mathbb{E}[\psi(Y) X].$$

□

Question 5.1.5: If there exists another random variable Y' such that $\sigma(Y) = \sigma(Y')$, prove that

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid Y'], \quad \text{a.s.}$$

From this, we know that the "correct" way to define a conditional expectation is through a σ-algebra but not through a random variable. In the next subsection, we will start from this observation and define the conditional expectation in a more general setting.

5.2 Case of General Random Variables

In this subsection, we use Proposition 5.1.4 as the way to define conditional expectations.

5.2.1 Integrable Random Variables

We extend the definition of conditional expectations to a more general setting. We have the following theorem, which is also the definition of conditional expectations.

Theorem 5.2.1 : *Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. There exists a unique random variable in $L^1(\Omega, \mathcal{B}, \mathbb{P})$, denoted $\mathbb{E}[X \mid \mathcal{B}]$, such that*

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[X \mathbb{1}_B] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{B}] \mathbb{1}_B]. \tag{5.1}$$

More generally speaking, for all $\mathcal{B}$-measurable bounded random variable Z, we have,

$$\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{B}] Z]. \tag{5.2}$$

If X is non negative, then $\mathbb{E}[X \mid \mathcal{B}]$ is also non-negative.

In the above theorem, Eq. (5.1) or the equivalent Eq. (5.2) is called the *characteristic property* of the conditional expectation $\mathbb{E}[X \mid \mathcal{B}]$. If the sub-$\sigma$-algebra $\mathcal{B}$ is generated by a random variable Y, then we can also write this conditional expectation as,

$$\mathbb{E}[X \mid \mathcal{B}] = \mathbb{E}[X \mid \sigma(Y)] = \mathbb{E}[X \mid Y].$$

Proof : We start with the uniqueness. Let X' and X'' be two random variables in $L^1(\Omega, \mathcal{B}, \mathbb{P})$ such that

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

Consider $B = \{X' > X''\} \in \mathcal{B}$, then,

$$\mathbb{E}\left[(X' - X'')\mathbb{1}_{\{X'>X''\}}\right] = 0,$$

which means that $X' \leqslant X''$ a.s., and similarly, $X'' \leqslant X'$ a.s.

Then we prove the existence. Suppose that X is a non-negative random variable and let $\mathbb{Q}$ be a finite measure defined on the measurable space $(\Omega, \mathcal{B})$,

$$\forall B \in \mathcal{B}, \quad \mathbb{Q}(B) = \mathbb{E}[X\mathbb{1}_B].$$

If we restrict the probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{B})$, then we have, $\mathbb{Q} \ll \mathbb{P}$. From the Radon–Nikodym theorem, we know that there exists a non-negative $\mathcal{B}$-measurable random variable $\widetilde{X}$ such that,

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[X\mathbb{1}_B] = \mathbb{Q}(B) = \mathbb{E}[\widetilde{X}\mathbb{1}_B].$$

Take $B = \Omega$, we have $\mathbb{E}[\widetilde{X}] = \mathbb{E}[X] < \infty$, so $\widetilde{X} \in L^1(\Omega, \mathcal{B}, \mathbb{P})$. Hence, we have checked that $\mathbb{E}[X \mid \mathcal{B}] := \widetilde{X}$ satisfies the condition Eq. (5.1) in the statement. If X is not non-negative, then we only need to consider,

$$\mathbb{E}[X \mid \mathcal{B}] = \mathbb{E}[X^+ \mid \mathcal{B}] - \mathbb{E}[X^- \mid \mathcal{B}].$$

Finally, to deduce Eq. (5.2) from Eq. (5.1), we use the fact that a measurable function can be approximated by simple functions (Proposition 1.2.14). □

Example 5.2.2 : Let $\Omega = (0,1]$, $\mathcal{A} = \mathcal{B}((0,1])$ and $\mathbb{P}(\mathrm{d}\omega) = \mathrm{d}\omega$. Given a positive integer n, let $\mathcal{B}$ be the σ-algebra generated by the intervals $\{(\frac{i-1}{n}, \frac{i}{n}) : 1 \leqslant i \leqslant n\}$. Given $f \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, then we have,

$$\mathbb{E}[f \mid \mathcal{B}] = \sum_{i=1}^{n} f_i \mathbb{1}_{(\frac{i-1}{n}, \frac{i}{n})},$$

where $f_i = n\int_{(i-1)/n}^{i/n} f(\omega)\,\mathrm{d}\omega$ is the average of the measurable function f on $(\frac{i-1}{n}, \frac{i}{n})$.

Proposition 5.2.3 : *The conditional expectation of an integrable random variable has the following properties.*

(1) *If X is $\mathcal{B}$-measurable, then $\mathbb{E}[X \mid \mathcal{B}] = X$.*

(2) $X \mapsto \mathbb{E}[X \mid \mathcal{B}]$ *is linear.*

(3) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{B}]] = \mathbb{E}[X]$.

(4) $|\mathbb{E}[X \mid \mathcal{B}]| \leqslant \mathbb{E}[|X| \mid \mathcal{B}]$ *a.s., so* $\mathbb{E}[|\mathbb{E}[X \mid \mathcal{B}]|] \leqslant \mathbb{E}[|X|]$.

(5) *If $X \geqslant X'$ a.s., then $\mathbb{E}[X \mid \mathcal{B}] \geqslant \mathbb{E}[X' \mid \mathcal{B}]$ a.s.*

Remark 5.2.4 : We note that if $\mathcal{B}$ is taken to be a trivial σ-algebra, $\mathcal{B} = \{\varnothing, \Omega\}$ for instance, then we recover properties of the expectation.

Proof :

(1) This comes from the uniqueness and the characteristic property in Theorem 5.2.1.

(2) In the same way, from the uniqueness and the characteristic property in Theorem 5.2.1, we note that for any $X, X' \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $\alpha, \alpha' \in \mathbb{R}$, the random variable $\alpha\,\mathbb{E}[X \mid \mathcal{B}] + \alpha'\,\mathbb{E}[X' \mid \mathcal{B}]$ and $\alpha X + \alpha' X'$ satisfy the same characteristic property Eq. (5.1).

(3) In Eq. (5.1), we take $B = \Omega$ to obtain what needs to be shown.

(4) If X is non-negative, then $\mathbb{E}[X \mid \mathcal{B}]$ is also non-negative, so

$$|\,\mathbb{E}[X \mid \mathcal{B}]| = |\,\mathbb{E}[X^+ \mid \mathcal{B}] - \mathbb{E}[X^- \mid \mathcal{B}]| \leqslant \mathbb{E}[X^+ \mid \mathcal{B}] + \mathbb{E}[X^- \mid \mathcal{B}] = \mathbb{E}[|X| \mid \mathcal{B}].$$

(5) It comes from the linearity.

$\square$

5.2.2 Non-negative Random Variables

Theorem 5.2.5 : *Let X be a non-negative random variable. Then, the random variable defined by*

$$\mathbb{E}[X \mid \mathcal{B}] = \lim_{n\to\infty} \uparrow \mathbb{E}[X \wedge n \mid \mathcal{B}], \quad \text{a.s.}$$

is also a non-negative random variable which can be described by the following property: for all non-negative $\mathcal{B}$-measurable random variable Z, we have,

$$\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{B}]Z]. \tag{5.3}$$

Remark 5.2.6 : In the case that X is also integrable, we may compare Eq. (5.3) and Eq. (5.2), then we realize that the resulting conditional expectations $\mathbb{E}[X \mid \mathcal{B}]$ are the same. Therefore, we also say that Eq. (5.3) is the *characteristic property* of the conditional expectation if X is non-negative.

Proof : The non-decreasing property on the right side can be obtained from (5) of Proposition 5.2.3. Let Z be a non-negative $\mathcal{B}$-measurable random variable. The monotone convergence theorem implies,

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{B}]Z] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[X \wedge n \mid \mathcal{B}](Z \wedge n)] = \lim_{n\to\infty} \mathbb{E}[(X \wedge n)(Z \wedge n)] = \mathbb{E}[XZ].$$

Then, we prove the uniqueness. Let X' and X'' be two non-negative $\mathcal{B}$-measurable random variables such that for all non-negative $\mathcal{B}$-measurable random variable Z, we have,

$$\mathbb{E}[X'Z] = \mathbb{E}[X''Z].$$

Given $a, b \in \mathbb{Q}_+$, $a < b$, let
$$Z = \mathbb{1}_{\{X' \leqslant a < b \leqslant X''\}}.$$
Then, we obtain,
$$a\,\mathbb{P}(X' \leqslant a < b \leqslant X'') \geqslant b\,\mathbb{P}(X' \leqslant a < b \leqslant X''),$$
implying $\mathbb{P}(X' \leqslant a < b \leqslant X'') = 0$, and,
$$\mathbb{P}\Big(\bigcup_{\substack{a,b \in \mathbb{Q}_+ \\ a<b}} \{X' \leqslant a < b \leqslant X''\} \Big) = 0.$$
This means that $X' \geqslant X''$ a.s., and by symmetry, we also have $X'' \geqslant X'$ a.s. The proof is complete. □

Question 5.2.7: Under the following conditions, construct a random variable X such that $X < \infty$ a.s. but $\mathbb{P}(\mathbb{E}[X \mid \mathcal{B}] = \infty) > 0$.

(1) $\mathcal{B}$ is a trivial σ-algebra: $\mathcal{B} = \{\varnothing, \Omega\}$.

(2) $\mathcal{B}$ is a more complicated σ-algebra.

Proposition 5.2.8 : *The conditional expectations of non-negative random variables have the following properties.*

(1) If X and X' are non-negative random variables and $a, b \geqslant 0$, then,
$$\mathbb{E}[aX + bX' \mid \mathcal{B}] = a\,\mathbb{E}[X \mid \mathcal{B}] + b\,\mathbb{E}[X' \mid \mathcal{B}] \quad \text{a.s.}$$

(2) If X is $\mathcal{B}$-measurable, then $\mathbb{E}[X \mid \mathcal{B}] = X$, a.s.

(3) If (X_n) is a non-decreasing sequence of non-negative random variables and $X = \lim \uparrow X_n$, then,
$$\mathbb{E}[X \mid \mathcal{B}] = \lim_{n\to\infty} \uparrow \mathbb{E}[X_n \mid \mathcal{B}] \quad \text{a.s.}$$

(4) If (X_n) is a sequence of non-negative random variables, then,
$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \mid \mathcal{B}\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{B}] \quad \text{a.s.}$$

(5) Let (X_n) be a sequence of integrable random variables that converges almost surely to X. Suppose that there exists a non-negative random variable Z such that for all n, $|X_n| \leqslant Z$ a.s. and $\mathbb{E}[Z] < \infty$, then,
$$\mathbb{E}[X \mid \mathcal{B}] = \lim_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{B}] \quad \text{a.s.}$$
Moreover, the above convergence also holds in L^1.

(6) If φ is a non-negative convex function and $X \in L^1$, then,
$$\mathbb{E}[\varphi(X) \mid \mathcal{B}] \geqslant \varphi(\mathbb{E}[X \mid \mathcal{B}]) \quad \text{a.s..}$$

Remark 5.2.9 : In theory, the definition of the conditional expectation does not depend on any set of zero measure, which means that we need to write "a.s." in the definition. However, in practice, we will omit this the most of the time.

Proof : We can show (1) and (2) using the characteristic property of the conditional expectation.

(3) First note that, if two random variables satisfy $X_1 \geqslant X_2 \geqslant 0$, then we have $\mathbb{E}[X_1 \,|\, \mathcal{B}] \geqslant \mathbb{E}[X_2 \,|\, \mathcal{B}]$. Let $X' := \lim \uparrow \mathbb{E}[X_n \,|\, \mathcal{B}]$, then X' is a $\mathcal{B}$-measurable random variable with values in $[0, \infty]$. For any $\mathcal{B}$-measurable non-negative random variable Z, we have,

$$\mathbb{E}[ZX'] = \lim \uparrow \mathbb{E}[Z\,\mathbb{E}[X_n \,|\, \mathcal{B}]] = \lim \mathbb{E}[ZX_n] = \mathbb{E}[ZX].$$

Hence, from the characteristic property of the conditional expectation, we have, $X' = \mathbb{E}[X \,|\, \mathcal{B}]$.

(4) Use (3), we obtain,

$$\begin{aligned}
\mathbb{E}\left[\liminf_{n\to\infty} X_n \,\middle|\, \mathcal{B}\right] &= \mathbb{E}\left[\lim_{k\to\infty} \uparrow \left(\inf_{n\geqslant k} X_n\right) \middle|\, \mathcal{B}\right] \\
&= \lim_{k\to\infty} \uparrow \mathbb{E}\left[\inf_{n\geqslant k} X_n \,\middle|\, \mathcal{B}\right] \\
&\leqslant \lim_{k\to\infty} \left(\inf_{n\geqslant k} \mathbb{E}[X_n \,|\, \mathcal{B}]\right) \\
&= \liminf \mathbb{E}[X_n \,|\, \mathcal{B}].
\end{aligned}$$

(5) Use the inequality in (4) twice and we obtain,

$$\begin{aligned}
\mathbb{E}[Z - X \,|\, \mathcal{B}] &= \mathbb{E}\left[\liminf_{n\to\infty}(Z - X_n) \,\middle|\, \mathcal{B}\right] \leqslant \mathbb{E}[Z \,|\, \mathcal{B}] - \limsup_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{B}], \\
\mathbb{E}[Z + X \,|\, \mathcal{B}] &= \mathbb{E}\left[\liminf_{n\to\infty}(Z + X_n) \,\middle|\, \mathcal{B}\right] \leqslant \mathbb{E}[Z \,|\, \mathcal{B}] + \liminf_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{B}].
\end{aligned}$$

Hence,

$$\mathbb{E}[X \,|\, \mathcal{B}] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{B}] \leqslant \limsup_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{B}] \leqslant \mathbb{E}[X \,|\, \mathcal{B}].$$

This means that $\mathbb{E}[X_n \,|\, \mathcal{B}]$ converges almost surely to $\mathbb{E}[X \,|\, \mathcal{B}]$. The convergence in L^1 can be seen as a consequence of the dominated convergence theorem since we have,

$$|\,\mathbb{E}[X_n \,|\, \mathcal{B}]| \leqslant \mathbb{E}[|X_n| \,|\, \mathcal{B}] \leqslant \mathbb{E}[Z \,|\, \mathcal{B}],$$

and $\mathbb{E}[\mathbb{E}[Z \,|\, \mathcal{B}]] = \mathbb{E}[Z] < \infty$.

(6) See the proof of Jensen's inequality (Theorem 1.3.4).

□

Remark 5.2.10 : Given a probability measure $\mathbb{P}$, its expectation $\mathbb{E}$ and probability $\mathbb{P}$ have the following relation: for all $A \in \mathcal{A}$, we have $\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]$. Similarly, we can define the conditional probability as follows: for all $A \in \mathcal{A}$, we have $\mathbb{P}(A \,|\, \mathcal{B}) := \mathbb{E}[\mathbb{1}_A \,|\, \mathcal{B}]$. We note that such a definition results in a conditional probability which is also a *random variable*.

5.2.3 Special Case: Square-Integrable Random Variables

When $X \in L^2(\Omega, \mathcal{B}, \mathbb{P})$ is a square-integrable random variable, the conditional expectation $\mathbb{E}[X \mid \mathcal{B}]$ has an important interpretation. First, note that the random variables in $L^2(\Omega, \mathcal{B}, \mathbb{P})$ are also in $L^2(\Omega, \mathcal{A}, \mathbb{P})$ and have at least a representant (代表) which is $\mathcal{B}$-measurable. Additionally, $L^2(\Omega, \mathcal{B}, \mathbb{P})$ is a *closed subspace* of $L^2(\Omega, \mathcal{A}, \mathbb{P})$.

Theorem 5.2.11 : *If $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, then $\mathbb{E}[X \mid \mathcal{B}]$ is the orthogonal projection of X on $L^2(\Omega, \mathcal{B}, \mathbb{P})$.*

Proof : Using (6) from Proposition 5.2.8, we have,

$$\mathbb{E}[X \mid \mathcal{B}]^2 \leqslant \mathbb{E}[X^2 \mid \mathcal{B}], \quad \text{a.s.},$$

implying $\mathbb{E}[\mathbb{E}[X \mid \mathcal{B}]^2] \leqslant \mathbb{E}[X^2] < \infty$, which means $\mathbb{E}[X \mid \mathcal{B}] \in L^2(\Omega, \mathcal{B}, \mathbb{P})$.

Moreover, for all bounded $\mathcal{B}$-measurable random variable Z, using the characteristic property, we have,

$$\mathbb{E}[Z(X - \mathbb{E}[X \mid \mathcal{B}])] = \mathbb{E}[ZX] - \mathbb{E}[Z\,\mathbb{E}[X \mid \mathcal{B}]] = 0.$$

Hence, $X - \mathbb{E}[X \mid \mathcal{B}]$ is orthogonal to all the bounded and $\mathcal{B}$-measurable random variables. By density, $X - \mathbb{E}[X \mid \mathcal{B}]$ is orthogonal to $L^2(\Omega, \mathcal{B}, \mathbb{P})$. The proof is complete. □

Remark 5.2.12 : According to Theorem 5.2.11 above, we have the two following consequences.

(1) If X is square-integrable, then $\mathbb{E}[X \mid \mathcal{B}]$ is the best approximation of X among all the $\mathcal{B}$-measurable random variables by the L^2 norm.

(2) We do not need to use the Radon–Nikodym theorem to construct the conditional expectation. Using the orthogonal projection, we construct the conditional expectation for all the square-integrable random variables then extend by density this definition to all the integrable random variables.