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,

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

第五章 條件期望值

在這個章節中，我們要定義條件期望值，並且探討其性質。直覺上來說，給定一個 $\mathcal{A}$ 可測的隨機變數 X，他對於一個子 σ 代數 $\mathcal{B}$ 的條件期望值，可以看作是在 $\mathcal{B}$ 可測的隨機變數中，最能夠「逼近」X 的隨機變數。在處理實際的應用問題中，例如對未來做預測，由於「當下」對「未來的隨機變數」擁有的只是片面資訊，透過條件期望值的數學概念，讓我們可以以數學的方式來理解這樣的問題。

我們會先從離散隨機變數的例子開始，給出條件期望值的「特徵性質」，並在第 5.2 節中，將他當作條件期望值公理化定義的方式，定義可積隨機變數以及非負隨機變數的條件期望值。稍後，在第 5.4 節中，我們也會看到如何計算條件期望值，其中第 5.4.3 小節中，高斯隨機變數的狀況是值得注意的特例。最後我們以第 5.5 節中的條件分佈結束此章節，替第七章中所探討的馬可夫鍊鋪路。

第一節 離散隨機變數

給定機率空間 $(\Omega, \mathcal{A}, \mathbb{P})$，令 $B \in \mathcal{A}$ 為機率為正的可測事件 $\mathbb{P}(B) > 0$。我們可以在可測空間 $(\Omega, \mathcal{A})$ 上定義一個新的機率測度，稱作已知事件 B 時的條件機率 (conditional probability) 測度，寫作

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

同理，給定隨機變數 X，若 X 非負或是 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$，則 X 在已知事件 B 時的條件期望值 (conditional expectation) 寫作

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

我們可以把上式定義理解做 X 在機率測度 $\mathbb{P}(\cdot \,|\, B)$ 之下的期望值，可以想像做「在 B 成立的情況下，X 的平均」。

前面我們定義了隨機變數 X 在已知可測事件情況下的條件期望值；下一步，我們要推展這樣的概念，並定義 X 在已知隨機變數 Y 時的條件期望值。首先，考慮可數集合 E 以及值域在 E 上的隨機變數 Y。令 $E' = \{y \in E : \mathbb{P}(Y = y) > 0\}$。對於所有 $y \in E'$ 以及任意隨機變數 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$，根據上面我們可以定義

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

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,

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

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

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

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

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

Moreover, we can also notice that $\mathbb{E}[X \,|\, Y]$ 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 $\mathbb{E}[|\,\mathbb{E}[X \,|\, Y]|] \leqslant \mathbb{E}[|X|]$, which implies $\mathbb{E}[X \,|\, Y] \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Additionally, for all bounded $\sigma(Y)$-measurable random variable Z, we have,*

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

定義 5.1.1 ： 令 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$。已知 Y，隨機變數 X 的條件期望值是個隨機變數，且定義做

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

其中函數 $\varphi : E \longrightarrow \mathbb{R}$ 滿足

$$\varphi(y) = \begin{cases} \mathbb{E}[X \,|\, Y = y], & \text{若 } y \in E', \\ 0, & \text{其他狀況.} \end{cases}$$

註解 5.1.2 ： 在上述定義中，φ 在 $E \backslash E'$ 上的值是不重要的，而且不會影響到 $\mathbb{E}[X \,|\, Y]$ 的定義，因為

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

因此我們若更改 φ 在 $E \backslash E'$ 上的定義，得到的條件期望值 $\mathbb{E}[X \,|\, Y]$ 除了在測度為零的集合之上以外，還是相等的。

現在我們可以注意到，在隨機變數之下的條件期望值，本身是個隨機變數，此隨機變數滿足這樣的條件：若我們知道 Y 的值，那麼此條件期望值會給出 X 在此情況下的平均值。換句話說，

$$\text{若 } Y(\omega) = y \in E'\text{，則 } \mathbb{E}[X \,|\, Y](\omega) = \mathbb{E}[X \,|\, Y = y]\text{。}$$

此外，我們可以注意到，由於 $\mathbb{E}[X \,|\, Y]$ 是個 Y 的函數，因此對於 $\sigma(Y)$ 來說是可測的。稍後我們也會看到，這是在所有 $\sigma(Y)$ 可測的隨機變數中，「最能」趨近 X 的隨機變數。

範例 5.1.3 ： 令 $\Omega = \{1, \ldots, 6\}$ 以及對於所有 $\omega \in \Omega$，定義 $\mathbb{P}(\{\omega\}) = \frac{1}{6}$。定義下列隨機變數：

$$Y(\omega) = \begin{cases} 1 & \text{若 } \omega \text{ 為奇數,} \\ 0 & \text{若 } \omega \text{ 為偶數.} \end{cases}$$

接著設 $X(\omega) = \omega$，則我們有

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

命題 5.1.4 ： 我們有 $\mathbb{E}[|\,\mathbb{E}[X \,|\, Y]|] \leqslant \mathbb{E}[|X|]$。從此得知，我們有 $\mathbb{E}[X \,|\, Y] \in L^1(\Omega, \mathcal{A}, \mathbb{P})$。此外，對於所有 $\sigma(Y)$ 可測的有界隨機變數 Z，我們有

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

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.

證明：根據條件期望值 $\mathbb{E}[X \mid Y]$ 的定義，我們有

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

針對命題第二部份的敘述，給定 $\sigma(Y)$ 可測的有界隨機變數 Z，由命題 2.2.6 ，我們知道存在有界可測函數 ψ 使得 $Z = \psi(Y)$，則

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

□

問題 5.1.5：若存在另一個隨機變數 Y' 使得 $\sigma(Y) = \sigma(Y')$，試證

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

由此可知，定義條件期望值的「正確」方式是針對 σ 代數來定義，而不是針對隨機變數來定義。在下面的章節中，我們將從這個觀察出發，定義在更一般情況下的條件期望值。

第二節　一般的隨機變數

這個小節中，我們將命題 5.1.4 中的性質當作定義條件期望值的方式。

第一小節　可積隨機變數

我們將上個章節中看到的離散條件期望值推廣到一般的情況下。我們有下列的定理，同時也是一般情況下條件期望值的定義。

定理 5.2.1 ：令 $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ 以及 $\mathcal{A}$ 的子 σ 代數 $\mathcal{B}$。存在唯一在 $L^1(\Omega, \mathcal{B}, \mathbb{P})$ 中的隨機變數，記作 $\mathbb{E}[X \mid \mathcal{B}]$ 使得

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

更一般來說，對於所有 $\mathcal{B}$ 可測的有界隨機變數 Z，我們有

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

若 X 為非負隨機變數，則 $\mathbb{E}[X \mid \mathcal{B}]$ 也非負。

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). $\square$

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 $(\frac{i-1}{n}, \frac{i}{n})$.

在上述定理中，式 (5.1) 或等價的式 (5.2) 稱作條件期望值 $\mathbb{E}[X \mid \mathcal{B}]$ 的特徵性質。若子 σ 代數 $\mathcal{B}$ 是由隨機變數 Y 所生成的，則我們也可以把此條件期望值記作

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

證明：我們先證明唯一性。令 X' 及 X'' 為兩個在 $L^1(\Omega, \mathcal{B}, \mathbb{P})$ 中的隨機變數使得

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

取 $B = \{X' > X''\} \in \mathcal{B}$，則

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

也就是說，$X' \leqslant X''$ a.s.，同理 $X'' \leqslant X'$ a.s.。

接著證明存在性。假設 X 是個非負的隨機變數，並令 $\mathbb{Q}$ 為定義在可測空間 $(\Omega, \mathcal{B})$ 上的有限測度：

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

若我們將機率測度 $\mathbb{P}$ 限制在可測空間 $(\Omega, \mathcal{B})$ 上，則我們有 $\mathbb{Q} \ll \mathbb{P}$。根據 Radon–Nikodym 定理，我們得知存在一個 $\mathcal{B}$ 可測的非負隨機變數 $\widetilde{X}$，使得

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

取 $B = \Omega$，我們有 $\mathbb{E}[\widetilde{X}] = \mathbb{E}[X] < \infty$，因此 $\widetilde{X} \in L^1(\Omega, \mathcal{B}, \mathbb{P})$。因此，我們驗證了 $\mathbb{E}[X \mid \mathcal{B}] := \widetilde{X}$ 滿足定理敘述要求的條件式 (5.1)。若 X 不是非負，則我們只需要取

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

最後，從式 (5.1) 推得式 (5.2)，用到的是可測函數可以被簡單函數所逼近的性質（命題 1.2.14）。 $\square$

範例 5.2.2 ： 令 $\Omega = (0,1]$，$\mathcal{A} = \mathcal{B}((0,1])$ 以及 $\mathbb{P}(\mathrm{d}\omega) = \mathrm{d}\omega$。給定正整數 n，令 $\mathcal{B}$ 是由區間 $\{(\frac{i-1}{n}, \frac{i}{n}) : 1 \leqslant i \leqslant n\}$ 所生成的 σ 代數。給定 $f \in L^1(\Omega, \mathcal{A}, \mathbb{P})$，則我們有

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

其中 $f_i = n \int_{(i-1)/n}^{i/n} f(\omega)\,\mathrm{d}\omega$ 是可測函數 f 在 $(\frac{i-1}{n}, \frac{i}{n})$ 上的平均。

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. □

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,

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

命題 5.2.3 ： 可積隨機變數的條件期望值有以下性質。

(1) 若 X 是 $\mathcal{B}$ 可測的，則 $\mathbb{E}[X \mid \mathcal{B}] = X$。

(2) $X \mapsto \mathbb{E}[X \mid \mathcal{B}]$ 是線性的。

(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.，因此 $\mathbb{E}[|\mathbb{E}[X \mid \mathcal{B}]|] \leqslant \mathbb{E}[|X|]$。

(5) 若 $X \geqslant X'$ a.s.，則 $\mathbb{E}[X \mid \mathcal{B}] \geqslant \mathbb{E}[X' \mid \mathcal{B}]$ a.s.。

註解 5.2.4 ： 我們注意到，若將 $\mathcal{B}$ 取做平凡 σ 代數，例如 $\mathcal{B} = \{\varnothing, \Omega\}$，則我們得到的是期望值的性質。

證明：

(1) 這可由定理 5.2.1 中的唯一性以及特徵性質得到。

(2) 同樣根據定理 5.2.1 中的唯一性以及特徵性質，我們注意到對於任意 $X, X' \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ 以及 $\alpha, \alpha' \in \mathbb{R}$，隨機變數 $\alpha\,\mathbb{E}[X \mid \mathcal{B}] + \alpha'\,\mathbb{E}[X' \mid \mathcal{B}]$ 以及 $\alpha X + \alpha' X'$ 滿足同樣的特徵性質式 (5.1)。

(3) 在式 (5.1) 中，我們取 $B = \Omega$ 便可以得到。

(4) 若 X 非負，則 $\mathbb{E}[X \mid \mathcal{B}]$ 也是非負，所以

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

(5) 根據線性性質，得證。 □

第二小節 非負隨機變數

定理 5.2.5 ： 令 X 為非負的隨機變數，則由

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

所定義的也是個非負隨機變數，而且是被下列性質所描述的：對於所有 $\mathcal{B}$ 可測的非負隨機變數 Z，我們有

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

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

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

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 $\mathbb{P}(\mathbb{E}[X\,|\,\mathcal{B}]=\infty)>0$.

(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,

$$\mathbb{E}[aX+bX'\,|\,\mathcal{B}] = a\,\mathbb{E}[X\,|\,\mathcal{B}] + b\,\mathbb{E}[X'\,|\,\mathcal{B}] \quad \text{a.s.}$$

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

(3) If (X_n) is a non-decreasing sequence of non-negative random variables and $X=\lim\uparrow X_n$, then,

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

註解 5.2.6 ： 在 X 也同時為可積的情況下，若將式 (5.3) 與式 (5.2) 比較，我們定義出來的條件期望值 $\mathbb{E}[X\,|\,\mathcal{B}]$ 是相同的。在非負隨機變數的情況下，我們也把式 (5.3) 稱作條件期望值的特徵性質。

證明： 極限右側的非遞減性可以從命題 5.2.3 中的 (5) 所得到。令 Z 為 $\mathcal{B}$ 可測的非負隨機變數，單調收斂定理告訴我們

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

接著我們證明唯一性。令 X' 以及 X'' 兩個 $\mathcal{B}$ 可測的非負隨機變數使得對於所有 $\mathcal{B}$ 可測的非負隨機變數 Z，我們有

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

給定 $a, b\in\mathbb{Q}_+$，$a<b$，令

$$Z = \mathbb{1}_{\{X'\leqslant a<b\leqslant X''\}}.$$

則我們可以得到

$$a\,\mathbb{P}(X'\leqslant a<b\leqslant X'') \geqslant b\,\mathbb{P}(X'\leqslant a<b\leqslant X''),$$

因此 $\mathbb{P}(X'\leqslant a<b\leqslant X'')=0$，所以

$$\mathbb{P}\Big(\bigcup_{\substack{a,b\in\mathbb{Q}_+\\ a<b}} \{X'\leqslant a<b\leqslant X''\}\Big) = 0.$$

這代表著 $X'\geqslant X''$ a.s.，根據對稱性，我們也有 $X''\geqslant X'$ a.s.，得證。 □

問題 5.2.7： 在下列情況中，構造隨機變數 X 使得 $X<\infty$ a.s. 但 $\mathbb{P}(\mathbb{E}[X\,|\,\mathcal{B}]=\infty)>0$。

(1) $\mathcal{B}$ 為平凡 σ 代數：$\mathcal{B}=\{\varnothing,\Omega\}$。

(2) $\mathcal{B}$ 為稍微複雜一點的 σ 代數。

命題 5.2.8 ： 非負隨機變數的條件期望值有下列性質。

(1) 若 X 以及 X' 為非負隨機變數且 $a, b\geqslant 0$，則

$$\mathbb{E}[aX+bX'\,|\,\mathcal{B}] = a\,\mathbb{E}[X\,|\,\mathcal{B}] + b\,\mathbb{E}[X'\,|\,\mathcal{B}] \quad \text{a.s.}$$

(2) 若 X 是 $\mathcal{B}$ 可測的，則 $\mathbb{E}[X\,|\,\mathcal{B}]=X$, a.s.。

(3) 若 (X_n) 是由非負隨機變數構成的遞增序列，且 $X=\lim\uparrow X_n$，則

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

(4) *If (X_n) is a sequence of non-negative random variables, then,*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \,\middle|\, \mathcal{B}\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{B}] \quad \text{a.s.}$$

(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,*

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

Moreover, the above convergence also holds in L^1.

(6) *If φ is a non-negative convex function and $X \in L^1$, then,*

$$\mathbb{E}[\varphi(X) \,|\, \mathcal{B}] \geqslant \varphi(\mathbb{E}[X \,|\, \mathcal{B}]) \quad \text{a.s..}$$

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,

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

(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,

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

(4) 若 (X_n) 是由非負隨機變數構成的序列，則

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \,\middle|\, \mathcal{B}\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{B}] \quad \text{a.s.}$$

(5) 令 (X_n) 是由可積隨機變數構成的序列，且假設其幾乎必然收斂至 X。假設存在非負隨機變數 Z 使得對於所有 n，$|X_n| \leqslant Z$ a.s. 且 $\mathbb{E}[Z] < \infty$，則

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

此外，上述收斂在 L^1 中也成立。

(6) 若 φ 是個非負的凸函數，且 $X \in L^1$，則

$$\mathbb{E}[\varphi(X) \,|\, \mathcal{B}] \geqslant \varphi(\mathbb{E}[X \,|\, \mathcal{B}]) \quad \text{a.s..}$$

註解 5.2.9 ： 照理說，由於條件期望值的定義不取決於任意機率測度為零的集合，因此在定義時得註明 a.s.，不過這樣的註記很多時候我們會省略。

證明：(1) 及 (2) 利用條件期望值得特徵性質而得證。

(3) 首先注意到，若兩隨機變數滿足 $X_1 \geqslant X_2 \geqslant 0$ 則我們有 $\mathbb{E}[X_1 \,|\, \mathcal{B}] \geqslant \mathbb{E}[X_2 \,|\, \mathcal{B}]$。令 $X' := \lim \uparrow \mathbb{E}[X_n \,|\, \mathcal{B}]$，則 X' 會是個 $\mathcal{B}$ 可測的隨機變數，且其值域為 $[0, \infty]$。對於所有 $\mathcal{B}$ 可測的非負隨機變數 Z，我們有：

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

因此根據條件期望值得特徵性質，我們有 $X' = \mathbb{E}[X \,|\, \mathcal{B}]$。

(4) 使用 (3)，我們可以得到

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

(5) 我們使用 (4) 中的不等式兩次：

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

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

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

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

(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 $\mathbb{P}(A \mid \mathcal{B}) := \mathbb{E}[\mathbb{1}_A \mid \mathcal{B}]$. 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. □

因此我們得到

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

也就是說，$\mathbb{E}[X_n \mid \mathcal{B}]$ 幾乎必然收斂至 $\mathbb{E}[X \mid \mathcal{B}]$。$L^1$ 空間中的收斂可以看作是勒貝格控制收斂定理的結果，因為我們有

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

而且 $\mathbb{E}[\mathbb{E}[Z \mid \mathcal{B}]] = \mathbb{E}[Z] < \infty$。

(6) 參照琴生不等式的證明（定理 1.3.4）。

□

註解 5.2.10 ： 給定機率測度 $\mathbb{P}$，其期望值 $\mathbb{E}$ 和 $\mathbb{P}$ 有下列關係：對於所有 $A \in \mathcal{A}$，我們有 $\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]$。同理，我們可以將條件機率定義為：對於所有 $A \in \mathcal{A}$，我們有 $\mathbb{P}(A \mid \mathcal{B}) := \mathbb{E}[\mathbb{1}_A \mid \mathcal{B}]$。我們注意到，這樣定義的條件機率會是個隨機變數。

第三小節 特例：平方可積的隨機變數

當 $X \in L^2(\Omega, \mathcal{B}, \mathbb{P})$ 是個平方可積的隨機變數時，$\mathbb{E}[X \mid \mathcal{B}]$ 有個重要的詮釋方式。首先，注意到 $L^2(\Omega, \mathcal{B}, \mathbb{P})$ 中的隨機變數，是在 $L^2(\Omega, \mathcal{A}, \mathbb{P})$ 中，且至少有個 $\mathcal{B}$ 可測代表 (representant) 的。此外，$L^2(\Omega, \mathcal{B}, \mathbb{P})$ 也是個在 $L^2(\Omega, \mathcal{A}, \mathbb{P})$ 中的封閉子空間。

定理 5.2.11 ： 若 $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$，則 $\mathbb{E}[X \mid \mathcal{B}]$ 是 X 在 $L^2(\Omega, \mathcal{B}, \mathbb{P})$ 上的正交投影。

證明： 利用命題 5.2.8 中的 (6)，我們有

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

因此 $\mathbb{E}[\mathbb{E}[X \mid \mathcal{B}]^2] \leqslant \mathbb{E}[X^2] < \infty$，也就是說，$\mathbb{E}[X \mid \mathcal{B}] \in L^2(\Omega, \mathcal{B}, \mathbb{P})$。

此外，對於所有 $\mathcal{B}$ 可測且有界的隨機變數 Z，利用條件期望值的特徵性質，我們可以得到

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

所以，$X - \mathbb{E}[X \mid \mathcal{B}]$ 和所有 $\mathcal{B}$ 可測且有界的隨機變數是正交的。利用稠密性，$X - \mathbb{E}[X \mid \mathcal{B}]$ 和 $L^2(\Omega, \mathcal{B}, \mathbb{P})$ 是正交的，因此得證。 □

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.

註解 5.2.12 ： 根據上述定理 5.2.11 ，我們有下列兩個結果：

(1) 若 X 為平方可積，則 $\mathbb{E}[X \mid \mathcal{B}]$ 是在所有 $\mathcal{B}$ 可測的隨機變數中，X 用 L^2 範數最好的逼近。

(2) 我們可以不必使用 Radon–Nikodym 定理來構造條件期望值。我們利用投影的性質，對所有平方可積的隨機變數來構造條件期望值，在利用稠密性來構造所有可積隨機變數的條件期望值。