

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

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

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，我們要推展這樣的觀念，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

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

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 $\varphi : E \rightarrow \mathbb{R}$ 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 $E \setminus E'$ 上的值是不重要的，而且不會影響到 $\mathbb{E}[X | Y]$ 的定義，因為

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

因此我們若更改 φ 在 $E \setminus E'$ 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，

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

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 $\Omega = \{1, \dots, 6\}$ 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

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

命題 5.1.4：我們有 $\mathbb{E}[\mathbb{E}[X | Y]] \leq \mathbb{E}[|X|]$ 。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

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

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

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

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

$$\mathbb{E}[\psi(Y) \mathbb{E}[X | Y]] = \sum_{y \in E} \psi(y) \mathbb{E}[X \mathbf{1}_{Y=y}] = \sum_{y \in E} \mathbb{E}[\psi(Y) X \mathbf{1}_{Y=y}] = \mathbb{E}[\psi(Y) X]. \quad \square$$

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

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

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 $\mathbb{E}[X | \mathcal{B}]$ 使得

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[X \mathbf{1}_B] = \mathbb{E}[\mathbb{E}[X | \mathcal{B}] \mathbf{1}_B]. \quad (5.1)$$

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

$$\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X | \mathcal{B}]Z]. \quad (5.2)$$

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

在上述定理中，式 (5.1) 或等價的式 (5.2) 稱作條件期望值 $\mathbb{E}[X | \mathcal{B}]$ 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

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

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

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

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

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

也就是說， $X' \leq X''$ a.s.，同理 $X'' \leq X'$ 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：

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

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 $\tilde{X}$ ，使得

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

取 $B = \Omega$ ，我們有 $\mathbb{E}[\tilde{X}] = \mathbb{E}[X] < \infty$ ，因此 $\tilde{X} \in L^1(\Omega, \mathcal{B}, \mathbb{P})$ 。因此，我們驗證了 $\mathbb{E}[X | \mathcal{B}] := \tilde{X}$ 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

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

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

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

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

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 $\mathbb{E}[X | \mathcal{B}] = X$ 。
- (2) $X \mapsto \mathbb{E}[X | \mathcal{B}]$ 是線性的。
- (3) $\mathbb{E}[\mathbb{E}[X | \mathcal{B}]] = \mathbb{E}[X]$ 。
- (4) $|\mathbb{E}[X | \mathcal{B}]| \leq \mathbb{E}[|X| | \mathcal{B}]$ a.s.，因此 $\mathbb{E}[|\mathbb{E}[X | \mathcal{B}]|] \leq \mathbb{E}[|X|]$ 。

(5) 若 $X \geq X'$ a.s.，則 $\mathbb{E}[X | \mathcal{B}] \geq \mathbb{E}[X' | \mathcal{B}]$ 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 $\mathcal{B} = \{\emptyset, \Omega\}$ ，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 $\alpha \mathbb{E}[X | \mathcal{B}] + \alpha' \mathbb{E}[X' | \mathcal{B}]$ 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 $\mathbb{E}[X | \mathcal{B}]$ 也是非負，所以

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

- (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

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

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

$$\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X | \mathcal{B}]Z]. \quad (5.3)$$

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

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

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

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

則我們可以得到

$$a \mathbb{P}(X' \leq a < b \leq X'') \geq b \mathbb{P}(X' \leq a < b \leq X''),$$

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

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

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

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： $\mathcal{B} = \{\emptyset, \Omega\}$ 。
- (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 $a, b \geq 0$ ，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

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

- (4) 若 (X_n) 是由非負隨機變數構成的序列，則

$$\mathbb{E}[\liminf_{n \rightarrow \infty} X_n | \mathcal{B}] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{B}] \quad \text{a.s.}$$

- (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 ， $|X_n| \leq Z$ a.s. 且 $\mathbb{E}[Z] < \infty$ ，則

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

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

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

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 $X_1 \geq X_2 \geq 0$ 則我們有 $\mathbb{E}[X_1 | \mathcal{B}] \geq \mathbb{E}[X_2 | \mathcal{B}]$ 。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

$$\begin{aligned} \mathbb{E}[\liminf_{n \rightarrow \infty} X_n | \mathcal{B}] &= \mathbb{E}\left[\lim_{k \rightarrow \infty} \uparrow \left(\inf_{n \geq k} X_n\right) \middle| \mathcal{B}\right] \\ &= \lim_{k \rightarrow \infty} \uparrow \mathbb{E}\left[\inf_{n \geq k} X_n \middle| \mathcal{B}\right] \\ &\leq \lim_{k \rightarrow \infty} \left(\inf_{n \geq k} \mathbb{E}[X_n | \mathcal{B}]\right) \\ &= \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{B}]. \end{aligned}$$

(5) 我們使用 (4) 中的不等式兩次：

$$\mathbb{E}[Z - X | \mathcal{B}] = \mathbb{E}[\liminf_{n \rightarrow \infty} (Z - X_n) | \mathcal{B}] \leq \mathbb{E}[Z | \mathcal{B}] - \limsup_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{B}],$$

$$\mathbb{E}[Z + X | \mathcal{B}] = \mathbb{E}[\liminf_{n \rightarrow \infty} (Z + X_n) | \mathcal{B}] \leq \mathbb{E}[Z | \mathcal{B}] + \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{B}].$$

因此我們得到

$$\mathbb{E}[X | \mathcal{B}] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{B}] \leq \limsup_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{B}] \leq \mathbb{E}[X | \mathcal{B}].$$

也就是說， $\mathbb{E}[X_n | \mathcal{B}]$ 幾乎必然收斂至 $\mathbb{E}[X | \mathcal{B}]$ 。 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

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

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

(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 $\mathbb{P}(A | \mathcal{B}) := \mathbb{E}[\mathbb{1}_A | \mathcal{B}]$ 。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， $\mathbb{E}[X | \mathcal{B}]$ 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 $\mathbb{E}[X | \mathcal{B}]$ 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

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

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

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

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

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

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 $\mathbb{E}[X | \mathcal{B}]$ 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。