

Lecture-03: Conditional Expectation

1 Conditional expectation

Consider a probability space (Ω, \mathcal{F}, P) , a random variable $X : \Omega \rightarrow \mathbb{R}$ defined on this probability space, and a nontrivial event $E \in \mathcal{F}$ such that $P(E) > 0$.

Definition 1.1. The conditional distribution of X conditioned on a nontrivial event E is defined for each $x \in \mathbb{R}$ as

$$F_{X|E}(x) \triangleq \frac{P(\{X \leq x\} \cap E)}{P(E)}.$$

Remark 1. We can verify that $F_{X|E} : \mathbb{R} \rightarrow [0, 1]$ is a distribution function for any nontrivial $E \in \mathcal{F}$.

Definition 1.2. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a Borel measurable function. The conditional expectation of random variable $g(X)$ given a nontrivial event E is defined as

$$\mathbb{E}[g(X) | E] \triangleq \int_{x \in \mathbb{R}} g(x) dF_{X|E}(x).$$

Example 1.3. Consider two random variables X, Y defined on the same probability space (Ω, \mathcal{F}, P) with the joint distribution $F_{X,Y}(x,y) = P(\{X \leq x, Y \leq y\})$. For each $y \in \mathbb{R}$, we define event $A_Y(y) \triangleq Y^{-1}(-\infty, y] \in \mathcal{F}$ such that $F_Y(y) = P(A_Y(y))$. Then, for each $y \in \mathbb{R}$ such that $P(A_Y(y)) > 0$, we can write the conditional distribution of X given the event $A_Y(y)$ as

$$F_{X|A_Y(y)}(x) = \frac{F_{X,Y}(x,y)}{F_Y(y)}.$$

The conditional expectation of X given the event $A_Y(y)$ is defined as

$$\mathbb{E}[X|A_Y(y)] = \int_{x \in \mathbb{R}} x dF_{X|A_Y(y)}(x) = \int_{x \in \mathbb{R}} x \frac{d_x F_{X,Y}(x,y)}{F_Y(y)} = \frac{1}{F_Y(y)} \int_{x \in \mathbb{R}} x \int_{z \leq y} dF_{X,Y}(x,z).$$

Example 1.4. Consider a random variable $X : \Omega \rightarrow \mathbb{R}$ and a simple random variable $Y : \Omega \rightarrow \mathcal{Y}$ defined on the same probability space. We observe that the conditional distribution of X given the nontrivial event $E_y = Y^{-1}\{y\}$ for $y \in \mathcal{Y}$ is

$$F_{X|E_y}(x) = \frac{P(\{X \leq x, Y = y\})}{P(E_y)} = \frac{1}{P(E_y)} \int_{t \leq x} d_t P(\{X \leq t, Y = y\}).$$

Therefore, the conditional expectation of X given the event E_y is

$$\mathbb{E}[X | E_y] = \mathbb{E}[X | Y = y] = \int_{x \in \mathbb{R}} x d_x F_{X|E_y}(x) = \frac{1}{P(E_y)} \int_{x \in \mathbb{R}} x d_x \mathbb{E}[\mathbb{1}_{\{X \leq x\}} \mathbb{1}_{E_y}] = \frac{\mathbb{E}[X \mathbb{1}_{E_y}]}{P(E_y)}.$$

Since $\mathbb{E}[X | E_y]$ is a scalar, we can write $\mathbb{E}[X \mathbb{1}_{E_y}] = \mathbb{E}[\mathbb{E}[X | E_y] \mathbb{1}_{E_y}]$.

Definition 1.5. The **conditional expectation** of X given an event subspace $\mathcal{E} \subseteq \mathcal{F}$ is denoted $\mathbb{E}[X | \mathcal{E}]$ and is a random variable $Z \triangleq \mathbb{E}[X | \mathcal{E}] : \Omega \rightarrow \mathbb{R}$ where

1. **measurability:** For each $B \in \mathcal{B}(\mathbb{R})$, we have $Z^{-1}(B) \in \mathcal{E}$, and
2. **orthogonality:** for each event $E \in \mathcal{E}$, we have $\mathbb{E}[X \mathbb{1}_E] = \mathbb{E}[Z \mathbb{1}_E]$, and
3. **integrability:** $\mathbb{E}|Z| < \infty$.

Proposition 1.6. *Conditional expectation is unique almost surely.*

Proof. Consider a random variable $X : \Omega \rightarrow \mathbb{R}$ defined on a probability space (Ω, \mathcal{F}, P) and a sub event space $\mathcal{E} \subseteq \mathcal{F}$. Let Z_1 and Z_2 be conditional expectations of X given \mathcal{E} . It suffices to show that $A_\epsilon \triangleq \{\omega \in \Omega : Z_1 - Z_2 > \epsilon\} \in \mathcal{E}$ and $B_\epsilon \triangleq \{\omega \in \Omega : Z_2 - Z_1 > \epsilon\} \in \mathcal{E}$ defined for each $\epsilon > 0$ has measure $P(A_\epsilon) = P(B_\epsilon) = 0$. From the definition of conditional expectation and linearity of expectation, we can write

$$0 \leq \epsilon P(A_\epsilon) < \mathbb{E}[(Z_1 - Z_2)\mathbb{1}_{A_\epsilon}] = \mathbb{E}[X\mathbb{1}_{A_\epsilon}] - \mathbb{E}[X\mathbb{1}_{A_\epsilon}] = 0.$$

Similarly, we can show that $P(B_\epsilon) = 0$, and the result follows. \square

Remark 2. Any random variable $Z : \Omega \rightarrow \mathbb{R}$ that satisfies the measurability, orthogonality, and integrability, is the conditional expectation of X given the sub-event space \mathcal{E} from the a.s. uniqueness of conditional expectations.

Remark 3. Intuitively, we think of the event subspace \mathcal{E} as describing the information we have. For each $A \in \mathcal{E}$, we know whether or not A has occurred. The conditional expectation $\mathbb{E}[X|\mathcal{E}]$ is the “best guess” of the value of X given the information \mathcal{E} .

Definition 1.7. Consider a random variable $X : \Omega \rightarrow \mathbb{R}$ and a random vector $Y : \Omega \rightarrow \mathbb{R}^n$ defined on the same probability space (Ω, \mathcal{F}, P) . The conditional expectation of X given Y is defined as

$$\mathbb{E}[X | Y] \triangleq \mathbb{E}[X | \sigma(Y)].$$

Proposition 1.8. *For two random variables $X, Y : \Omega \rightarrow \mathbb{R}$ defined on the same probability space (Ω, \mathcal{F}, P) , the conditional expectation $\mathbb{E}[X | Y]$ is a function of Y .*

Proof. We denote the conditional expectation $\mathbb{E}[X | Y]$ by a $\sigma(Y)$ -measurable random variable $Z : \Omega \rightarrow \mathbb{R}$. It suffices to show that for any $y \in \mathbb{R}$, the conditional expectation $Z(\omega)$ remains constant on the set of outcomes $\omega \in Y^{-1}\{y\}$. First, we show that for any event $A \in \sigma(Y)$, either $Y^{-1}\{y\} \subseteq A$ or $A \cap Y^{-1}\{y\} = \emptyset$. This follows from the fact that either $y \in A$ or $y \notin A$. Next, we suppose that there exists a $y \in \mathbb{R}$ and $\omega_1, \omega_2 \in Y^{-1}\{y\}$ such that $Z(\omega_1) \neq Z(\omega_2)$. It follows that there exists an event $B \triangleq Z^{-1}\{Z(\omega_1)\} \in \sigma(Z)$ such that $\omega_1 \in B$ and $\omega_2 \notin B$. Since Z is $\sigma(Y)$ -measurable, it follows that $B \in \sigma(Z) \subseteq \sigma(Y)$. This leads to a contradiction. \square

Proposition 1.9. *Let X, Y be random variables on the probability space (Ω, \mathcal{F}, P) such that $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$. Let \mathcal{G} and \mathcal{H} be sub-event spaces of \mathcal{F} . Then*

1. **linearity:** $\mathbb{E}[\alpha X + \beta Y | \mathcal{G}] = \alpha \mathbb{E}[X | \mathcal{G}] + \beta \mathbb{E}[Y | \mathcal{G}]$, a.s.
2. **monotonicity:** If $X \leq Y$ a.s., then $\mathbb{E}[X | \mathcal{G}] \leq \mathbb{E}[Y | \mathcal{G}]$, a.s.
3. **identity:** If X is \mathcal{G} -measurable and $\mathbb{E}|X| < \infty$, then $X = \mathbb{E}[X | \mathcal{G}]$ a.s. In particular, $c = \mathbb{E}[c | \mathcal{G}]$, for any constant $c \in \mathbb{R}$.
4. **conditional Jensen's inequality:** If $\psi : \mathbb{R} \rightarrow \mathbb{R}$ is convex and $\mathbb{E}|\psi(X)| < \infty$, then $\mathbb{E}[\psi(X) | \mathcal{G}] \geq \psi(\mathbb{E}[X | \mathcal{G}])$, a.s.
5. **pulling out what's known:** If Y is \mathcal{G} -measurable and $\mathbb{E}|XY| < \infty$, then $\mathbb{E}[XY | \mathcal{G}] = Y\mathbb{E}[X | \mathcal{G}]$, a.s.
6. **L^2 -projection:** If $\mathbb{E}|X|^2 < \infty$, then $\zeta^* = \mathbb{E}[X | \mathcal{G}]$ minimizes $\mathbb{E}[(X - \zeta)^2]$ over all \mathcal{G} -measurable random variables ζ such that $\mathbb{E}|\zeta|^2 < \infty$.
7. **tower property:** If $\mathcal{H} \subseteq \mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{H}] = \mathbb{E}[X | \mathcal{H}]$, a.s..
8. **irrelevance of independent information:** If \mathcal{H} is independent of $\sigma(\mathcal{G}, \sigma(X))$ then $\mathbb{E}[X | \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X | \mathcal{G}]$, a.s. In particular, if X is independent of \mathcal{H} , then $\mathbb{E}[X | \mathcal{H}] = \mathbb{E}[X]$ a.s.

Proof. Let X, Y be random variables on the probability space (Ω, \mathcal{F}, P) such that $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$. Let \mathcal{G} and \mathcal{H} be event spaces such that $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$.

1. **linearity:** Let $Z \triangleq \alpha \mathbb{E}[X | \mathcal{G}] + \beta \mathbb{E}[Y | \mathcal{G}]$, then since $\mathbb{E}[X | \mathcal{G}], \mathbb{E}[Y | \mathcal{G}]$ are \mathcal{G} -measurable, it follows that their linear combination Z is also \mathcal{G} -measurable. The integrability follows from the following triangle inequality and the monotonicity of expectation

$$|Z| \leq |\alpha| |\mathbb{E}[X | \mathcal{G}]| + |\beta| |\mathbb{E}[Y | \mathcal{G}]|.$$

Further, for any event $F \in \mathcal{G}$, from the linearity of expectation and definition of conditional expectation, we have

$$\mathbb{E}[Z\mathbb{1}_G] = \alpha \mathbb{E}[\mathbb{E}[X | \mathcal{G}]\mathbb{1}_G] + \beta \mathbb{E}[\mathbb{E}[Y | \mathcal{G}]\mathbb{1}_G] = \mathbb{E}[(\alpha X + \beta Y)\mathbb{1}_G].$$

2. **monotonicity:** Let $\epsilon > 0$ and define $A_\epsilon \triangleq \{\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[Y | \mathcal{G}] > \epsilon\} \in \mathcal{G}$. Then from the definition of conditional expectation, we have

$$0 \leq \mathbb{E}[(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[Y | \mathcal{G}]) \mathbb{1}_{A_\epsilon}] = \mathbb{E}[(X - Y) \mathbb{1}_{A_\epsilon}] \leq 0.$$

Thus, we obtain that $P(A_\epsilon) = 0$ for all $\epsilon > 0$. Taking limit $\epsilon \downarrow 0$, we get $0 = \lim_{\epsilon \downarrow 0} P(A_\epsilon) = P(\lim_{\epsilon \downarrow 0} A_\epsilon) = P(A_0)$.

3. **identity:** It follows from the definition that X satisfies all three conditions for conditional expectation. The event space generated by any constant function is the trivial event space $\{\emptyset, \Omega\} \subseteq \mathcal{G}$ for any event space. Hence, $\mathbb{E}[c | \mathcal{G}] = c$.
4. **conditional Jensen's inequality:** We will use the fact that a convex function can always be represented by the supremum of a family of affine functions. Accordingly, we will assume for a convex function $\psi : \mathbb{R} \rightarrow \mathbb{R}$, we have linear functions $\phi_i : \mathbb{R} \rightarrow \mathbb{R}$ and constants $c_i \in \mathbb{R}$ for all $i \in I$ such that $\psi = \sup_{i \in I} (\phi_i + c_i)$. For each $i \in I$, we have $\phi_i(\mathbb{E}[X | \mathcal{G}]) + c_i = \mathbb{E}[\phi_i(X) | \mathcal{G}] + c_i \leq \mathbb{E}[\psi(X) | \mathcal{G}]$ from the linearity and monotonicity of conditional expectation. It follows that

$$\psi(\mathbb{E}[X | \mathcal{G}]) = \sup_{i \in I} (\phi_i(\mathbb{E}[X | \mathcal{G}]) + c_i) \leq \mathbb{E}[\psi(X) | \mathcal{G}].$$

5. **pulling out what's known:** Let Y be \mathcal{G} -measurable and $\mathbb{E}|XY| < \infty$. Since Y is given to be \mathcal{G} -measurable, conditional expectation $\mathbb{E}[X | \mathcal{G}]$ is \mathcal{G} -measurable by definition, and product function is Borel measurable, it follows that $Y\mathbb{E}[X | \mathcal{G}]$ is \mathcal{G} -measurable. It suffices to show that $\mathbb{E}[XY \mathbb{1}_G] = \mathbb{E}[Y\mathbb{E}[X | \mathcal{G}] \mathbb{1}_G]$ for all events $G \in \mathcal{G}$ and $\mathbb{E}|Y\mathbb{E}[X | \mathcal{G}]| < \infty$, when Y is a simple \mathcal{G} -measurable random variable such that $\mathbb{E}|XY| < \infty$. It follows that, we can write $Y = \sum_{y \in \mathcal{Y}} y \mathbb{1}_{E_y}$ for finite \mathcal{Y} and $E_y \triangleq Y^{-1}\{y\} \in \mathcal{G}$ for all $y \in \mathcal{Y}$. From the definition of conditional expectation and linearity, we obtain for any $G \in \mathcal{G}$

$$\mathbb{E}[Y\mathbb{E}[X | \mathcal{G}] \mathbb{1}_G] = \sum_{y \in \mathcal{Y}} y \mathbb{E}[\mathbb{1}_{G \cap E_y} \mathbb{E}[X | \mathcal{G}]] = \sum_{y \in \mathcal{Y}} y \mathbb{E}[X \mathbb{1}_{G \cap E_y}] = \mathbb{E}[X \sum_{y \in \mathcal{Y}} y \mathbb{1}_{G \cap E_y}] = \mathbb{E}[XY \mathbb{1}_G].$$

Conditional Jensen's inequality applied to convex function $|\cdot| : \mathbb{R} \rightarrow \mathbb{R}_+$, we get $|\mathbb{E}[X | \mathcal{G}]| \leq \mathbb{E}[|X| | \mathcal{G}]$. Therefore,

$$\mathbb{E}[|Y| |\mathbb{E}[X | \mathcal{G}]|] = \sum_{y \in \mathcal{Y}} |y| \mathbb{E}[|\mathbb{E}[X | \mathcal{G}]| \mathbb{1}_{E_y}] \leq \sum_{y \in \mathcal{Y}} |y| \mathbb{E}[|X| \mathbb{1}_{E_y}] = \mathbb{E}|XY|.$$

6. **L^2 -projection:** We define $L^2(\mathcal{G}) \triangleq \{\zeta \text{ a } \mathcal{G} \text{ measurable random variable} : \mathbb{E}\zeta^2 < \infty\}$. From the conditional Jensen's inequality applied to convex function $(\cdot)^2 : \mathbb{R} \rightarrow \mathbb{R}_+$, we get that $\mathbb{E}(\mathbb{E}[X | \mathcal{G}])^2 \leq \mathbb{E}[X^2 | \mathcal{G}]$. Since $X \in L^2$, it follows that $X^2 \in L^1$ and hence $\mathbb{E}[X | \mathcal{G}] \in L^2$. It follows that $\zeta^* \triangleq \mathbb{E}[X | \mathcal{G}] \in L^2(\mathcal{G})$ from the definition of conditional expectation. We first show that $X - \zeta^*$ is uncorrelated with all $\zeta \in L^2(\mathcal{G})$. Towards this end, we let $\zeta \in L^2(\mathcal{G})$ and observe that

$$\mathbb{E}[(X - \zeta^*)\zeta] = \mathbb{E}[\zeta X] - \mathbb{E}[\zeta \mathbb{E}[X | \mathcal{G}]] = \mathbb{E}[\zeta X] - \mathbb{E}[\mathbb{E}[\zeta X | \mathcal{G}]] = 0.$$

The above equality follows from the linearity of expectation, the \mathcal{G} -measurability of ζ , and the definition of conditional expectation. Since $\zeta^* \in L^2(\mathcal{G})$, we have $(\zeta - \zeta^*) \in L^2(\mathcal{G})$. Therefore, $\mathbb{E}[(X - \zeta^*)(\zeta - \zeta^*)] = 0$. For any $\zeta \in L^2(\mathcal{G})$, we can write from the linearity of expectation

$$\mathbb{E}(X - \zeta)^2 = \mathbb{E}(X - \zeta^*)^2 + \mathbb{E}(\zeta - \zeta^*)^2 - 2\mathbb{E}(X - \zeta^*)(\zeta - \zeta^*) \geq \mathbb{E}(X - \zeta^*)^2.$$

7. **tower property:** Measurability follows from the definition of conditional expectation, since $\mathbb{E}[X | \mathcal{H}]$ is \mathcal{H} measurable. Integrability follows from the application of conditional Jensen's inequality to convex function $|\cdot| : \mathbb{R} \rightarrow \mathbb{R}_+$ to get $|\mathbb{E}[X | \mathcal{H}]| \leq \mathbb{E}[|X| | \mathcal{H}]$, which implies $\mathbb{E}|\mathbb{E}[X | \mathcal{H}]| \leq \mathbb{E}|X| < \infty$. Orthogonality follows from the definition of conditional expectation, since for any $H \in \mathcal{H} \subseteq \mathcal{G}$, we have

$$\mathbb{E}[\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{H}] \mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}] \mathbb{1}_H] = \mathbb{E}[X \mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X | \mathcal{H}] \mathbb{1}_H].$$

8. **irrelevance of independent information:** Measurability follows from the definition of conditional expectation and the definition of $\sigma(\mathcal{G}, \mathcal{H})$. Since $\mathbb{E}[X | \mathcal{G}]$ is \mathcal{G} -measurable, it is $\sigma(\mathcal{G}, \mathcal{H})$ measurable. Integrability follows from the conditional Jensen's inequality applied to convex function $|\cdot| : \mathbb{R} \rightarrow \mathbb{R}_+$ to get $|\mathbb{E}[X | \mathcal{G}]| \leq \mathbb{E}[|X| | \mathcal{G}]$, which implies that $\mathbb{E}|\mathbb{E}[X | \mathcal{G}]| \leq \mathbb{E}|X| < \infty$. Orthogonality follows from the fact that it suffices to show for events $A = G \cap H \in \sigma(\mathcal{G}, \mathcal{H})$ where $G \in \mathcal{G}$ and $H \in \mathcal{H}$. In this case,

$$\mathbb{E}[\mathbb{E}[X | \mathcal{G}] \mathbb{1}_{G \cap H}] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}] \mathbb{1}_G \mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}] \mathbb{1}_G] \mathbb{E}[\mathbb{1}_H] = \mathbb{E}[X \mathbb{1}_G] \mathbb{E}[\mathbb{1}_H] = \mathbb{E}[X \mathbb{1}_{G \cap H}].$$

□

Example 1.10 (Conditioning on simple random variables). Let X and Y be random variables defined on the probability space (Ω, \mathcal{F}, P) , where $Y = \sum_{y \in \mathcal{Y}} y \mathbb{1}_{E_y}$ is simple with finite \mathcal{Y} , $E_y \triangleq Y^{-1}\{y\} \in \mathcal{F}$ for all $y \in \mathcal{Y}$ are mutually disjoint, and $p_y \triangleq P(E_y) > 0$ for all $y \in \mathcal{Y}$. Then, we observe that

$$\mathbb{E}[X|Y] = \sum_{y \in \mathcal{Y}} \mathbb{E}[X | E_y] \mathbb{1}_{E_y} \text{ a.s.}$$

To show this, we will use the almost sure uniqueness of conditional expectation that satisfies three properties in the definition. For measurability, we observe that $\sigma(Y) = \sigma(E_y : y \in \mathcal{Y})$, and RHS is a simple $\sigma(Y)$ -measurable random variable. For integrability, we observe that

$$\mathbb{E} \left| \sum_{y \in \mathcal{Y}} \mathbb{E}[X | E_y] \mathbb{1}_{E_y} \right| \leq \sum_{y \in \mathcal{Y}} |\mathbb{E}[X | Y]| P(E_y).$$

Thus, integrability follows from the finiteness of $|\mathbb{E}[X | E_y]|$. For orthogonality, we observe that any $G \in \sigma(Y) = \cup_{y \in F} E_y$ for some finite subset $F \subseteq \mathcal{Y}$. Further, we observe that $\mathbb{E}[X \mathbb{1}_{E_y}] = \mathbb{E}[X | E_y] P(E_y)$. Therefore, we have

$$\mathbb{E} \left[\sum_{z \in F} \sum_{y \in \mathcal{Y}} \mathbb{E}[X | E_y] \mathbb{1}_{E_y} \mathbb{1}_{E_z} \right] = \mathbb{E} \left[\sum_{z \in F} \mathbb{E}[X | E_z] \mathbb{1}_{E_z} \right] = \mathbb{E}[X \mathbb{1}_G].$$

Example 1.11 (Conditioning on simple random variables). Consider two random variables X, Y defined on the same probability space (Ω, \mathcal{F}, P) , where Y is a simple random variable such that $\mathcal{Y} \subseteq \mathbb{R}$ is finite alphabet, $E_y \triangleq Y^{-1}(\{y\}) \in \sigma(Y) \subseteq \mathcal{F}$, and $p_y \triangleq P(E_y) > 0$. Thus, we can write

$$Y = \sum_{y \in \mathcal{Y}} y \mathbb{1}_{E_y}.$$

The collection $(E_y \in \mathcal{F} : y \in \mathcal{Y})$ forms a finite partition of the outcome space Ω and generates $\sigma(Y) = \{\cup_{y \in F} E_y \in \mathcal{F} : F \subseteq \mathcal{Y}\}$. For an event space $\mathcal{E} \subseteq \mathcal{F}$, we claim

$$\mathbb{E}[X | \sigma(\mathcal{E}, Y)] = \sum_{y \in \mathcal{Y}} \mathbb{E}[X | \sigma(\mathcal{E}, E_y)] \mathbb{1}_{E_y} \text{ a.s.}$$

We will show this by uniqueness of conditional expectation that satisfies the following three properties. First, we verify that RHS is $\sigma(\mathcal{E}, Y)$ measurable, which follows from the definition since $\mathbb{E}[X | \sigma(\mathcal{E}, E_y)] \in \sigma(\mathcal{E}, E_y) \subseteq \sigma(\mathcal{E}, Y)$. Second, it follows from the triangular and conditional Jensen's inequality, that

$$\mathbb{E} \left| \sum_{y \in \mathcal{Y}} \mathbb{E}[X | \sigma(\mathcal{E}, E_y)] \mathbb{1}_{E_y} \right| \leq \sum_{y \in \mathcal{Y}} \mathbb{E}[\mathbb{E}[|X| \mathbb{1}_{E_y} | \sigma(\mathcal{E}, E_y)]] \leq \mathbb{E}|X|.$$

It suffices to show that for any $A \in \mathcal{E}$ and $z \in \mathcal{Y}$, we have $\mathbb{E}[\sum_{y \in \mathcal{Y}} \mathbb{E}[X | \sigma(\mathcal{E}, E_y)] \mathbb{1}_{E_y} \mathbb{1}_A \mathbb{1}_{E_z}] = \mathbb{E}[X \mathbb{1}_A \mathbb{1}_{E_z}]$. To this end, we observe that LHS of above equation is equal to

$$\mathbb{E}[\mathbb{E}[X \mathbb{1}_{A \cap E_z} | \sigma(\mathcal{E}, E_z)]] = \mathbb{E}[X \mathbb{1}_{A \cap E_z}].$$