Given a Hilbert space H (like $L^2(\Omega,\mathcal{F},\mathbb{P})$), and a closed subspace $K\subseteq H$, we have the orthogonal projection:

$$P_K : H \twoheadrightarrow K$$

a linear transformation, with $\|P_K(X)\| \le \|X\|$, characterized by

- $P_K(X)$ is the unique $Z\in K$ minimizing $\|X-Z\|$
- $P_K(X)$ is the unique $Z\in K$ s.t. $X-Z\perp K$

$$\langle X-P_K(X), Y\rangle = 0 \quad \forall Y\in K.$$

↳ I.e. $\langle P_K(X),Y\rangle = \langle X,Y\rangle \quad \forall Y\in K$

$$\mathbb{E}[P_K(X)Y] = \mathbb{E}[XY] \qquad \text{if } H=L^2(\Omega,\mathcal{F},P)$$

Recall: if $\mathcal{G}=\sigma(\{A_n\}_{n=1}^{\infty})$ for a partition $\{A_n\}_{n=1}^{\infty}$ of Ω, then

$$\mathbb{E}[\mathbb{E}_{\mathcal{G}}[X]Y] = \mathbb{E}[XY] \qquad \forall\, Y\in L^2(\Omega,\mathcal{G},\mathbb{P})$$

Prop: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-field. Then $L^2(\Omega, \mathcal{G}, \mathbb{P}) \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a closed subspace.

Pf. We're identifying $L^2(\Omega, \mathcal{G}, \mathbb{P}) = \{X \in L^2(\Omega, \mathcal{F}, \mathbb{P}) : X \text{ is } \mathcal{G}\text{-measurable}\}$

Subspace: X, Y $\mathcal{G}$-meas. $\alpha, \beta \in \mathbb{R}$ $\alpha X + \beta Y$ is $\mathcal{G}$-measurable.

$$\|\alpha X + \beta Y\|_{L^2} \leq |\alpha| \|X\|_{L^2} + |\beta| \|Y\|_{L^2} < \infty$$

Closed: If X_n is $\mathcal{G}$-meas, $X_n \to X$ in $L^2(\Omega, \mathcal{F}, \mathbb{P})$

$\exists$ subseq. $X_{n_k} \to X$ a.s.

$\Rightarrow X$ is $\mathcal{G}$-meas.

$\therefore X \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. ///

Def: If $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ is any sub-σ-field, the conditional expectation $E_{\mathcal{G}}[X]$ is the random variable $= P_{L^2(\Omega, \mathcal{G}, \mathbb{P})}(X)$.

E.g. $\Omega = \{1,2,3\}$, $\mathcal{F} = 2^\Omega$

$$\mathcal{G} = \sigma(\{1,2\}, \{3\})$$

$$= \{\emptyset, \{1,2\}, \{3\}, \Omega\}$$

If X is $\mathcal{G}$-measurable, then $X^{-1}\{t\} \in$ for each $t \in \mathbb{R}$. In particular, with $t = X(1)$

$$\left.\begin{array}{l} 1 \in X^{-1}(X(1)) = \underset{X(1)=X(2)}{\{1,2\}} \text{ or } \underset{X(1)=X(2)=X(3)}{\{1,2,3\}} \end{array}\right\} \therefore X(1) = X(2)$$

$$\begin{tikzcd} u_1 \arrow[d] & u_2 \arrow[d] \\ \mathbb{1}_{\{1,2\}} \arrow[d] & \mathbb{1}_{\{3\}} \arrow[d] \\ & \end{tikzcd}$$

In fact, for any $\mathbb{P}$, $L^2(\Omega, \mathcal{F}, \mathbb{P}) = \mathbb{R}^{\{1,2,3\}} \cong \mathbb{R}^3 \supsetneq \operatorname{Span}\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$

$$\shortparallel$$

$$L^2(\Omega, \mathcal{G}, \mathbb{P}) = \{X \in \mathbb{R}^{\{1,2,3\}} : X(1) = X(2)\} \cong \begin{bmatrix}x\\x\\y\end{bmatrix}$$

Take $\mathbb{P} = \operatorname{Unif}\{1,2,3\}$. Then $\mathbb{E}[XY] = \frac{1}{3}(X(1)Y(1) + X(2)Y(2) + X(3)Y(3))$

$$\mathbb{E}_{\mathcal{G}}[X] = \frac{X \cdot u_1}{\|u_1\|^2} u_1 + \frac{X \cdot u_2}{\|u_2\|^2} u_2 = \frac{\mathbb{E}[X \mathbb{1}_{\{1,2\}}]}{\mathbb{P}\{1,2\}} \mathbb{1}_{\{1,2\}} + \frac{\mathbb{E}[X \mathbb{1}_{\{3\}}]}{\mathbb{P}\{3\}} \mathbb{1}_{\{3\}}$$

$\frac{2}{3} = \mathbb{P}\{1,2\}$ $\qquad$ $\frac{1}{3} = \mathbb{P}\{3\}$

What is $\mathbb{E}_{\mathcal{G}}[X]$? It is the $\mathcal{G}$-measurable r.v. that is closest to X

$$\| X - \mathbb{E}_{\mathcal{G}}[X] \|_{L^2} = \min_{Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \| X - Y \| .$$

I.e. it is the best guess for X, using only the information in $\mathcal{G}$.

Question: does it only make sense for L^2?

Lemma: If $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathbb{E}[|\mathbb{E}_{\mathcal{G}}[X]|] \leq \mathbb{E}[|X|]$.

Pf: Set $Y = \mathbb{E}_{\mathcal{G}}[X]$. Take $Z = \operatorname{sgn} Y = \begin{Bmatrix} 1, & Y > 0 \\ 0, & Y = 0 \\ -1, & Y < 0 \end{Bmatrix} = \mathbb{1}_{Y \neq 0} \frac{Y}{|Y|}$ ← $\mathcal{G}$-meas. b/c Y is $\mathcal{G}$-meas.

bounded ∴ L^2

,, K

∴ $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$, and so

$$\mathbb{E}[|Y|] = \mathbb{E}[\underset{\underset{P_K(X)}{\nearrow}}{Y} \cdot \operatorname{sgn} \underset{\underset{K}{\nwarrow}}{Y}] = \mathbb{E}[X \operatorname{sgn} Y] \leq \mathbb{E}[|X \operatorname{sgn} Y|] \leq \mathbb{E}[|X|].$$

///

The lemma shows that $\|E_{\mathcal{G}}[X]\|_{L^1} \leq \|X\|_{L^1}$ $\forall X \in L^2$.

I.e. if we equip $L^2(\Omega,\mathcal{F},\mathbb{P})$ with the L^1-norm,

$E_{\mathcal{G}}$ is still Lipschitz, .

Note that $L^2 \subseteq L^1$ is dense: given $X \in L^1$,

$X\mathbb{1}_{|X|\leq n}$ is bdd and ∴ in L^2, and $\|X - X\mathbb{1}_{|X|\geq n}\|_{L^1} = \mathbb{E}[|X|\mathbb{1}_{|X|>n}] \xrightarrow[n\to\infty]{\text{DCT}} 0$

Def: If $X \in L^1(\Omega,\mathcal{F},\mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$, define $E_{\mathcal{G}}[X]$ as follows:

↳ Take any sequence $X_n \in L^2(\Omega,\mathcal{F},\mathbb{P})$ s.t. $\|X_n - X\|_{L^1} \to 0$

↳ Define $E_{\mathcal{G}}[X] := L^1\text{-}\lim_{n\to\infty} E_{\mathcal{G}}[X_n] = L^1\text{-}\lim_{n\to\infty} P_{L^2(\Omega,\mathcal{G},\mathbb{P})}(X_n)$

- Exists: $\|E_{\mathcal{G}}[X_n] - E_{\mathcal{G}}[X_m]\|_{L^1} = \|E_{\mathcal{G}}[X_n - X_m]\|_{L^1} \leq \|X_n - X_m\|_{L^1} \to 0$ as $n,m \to \infty$.

  ∴ $\{E_{\mathcal{G}}[X_n]\}_{n\geq 1}^{\infty}$ is L^1-Cauchy. ∴ exists.

- Well-defined: if $X_n, Y_n \to X$ in L^1, then

  $\|E_{\mathcal{G}}[X_n] - E_{\mathcal{G}}[Y_n]\|_{L^1} = \|E_{\mathcal{G}}[X_n - Y_n]\|_{L^1} \leq \|X_n - Y_n\|_{L^1} \to 0$ as $n \to \infty$.

Prop (Averaging property / characterization)

For $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$, $\mathbb{E}_{\mathcal{G}}[X]$ is the unique $L^1(\Omega, \mathcal{G}, \mathbb{P})$ random variable with the property:

$$(\star) \qquad \mathbb{E}\big[\mathbb{E}_{\mathcal{G}}[X]\,Y\big] = \mathbb{E}[XY] \qquad \forall\, Y \in B(\Omega, \mathcal{G}) .$$

Pf. If $X \in L^2$, $(\star)$ holds by defⁿ of orth. proj. (b/c $B(\Omega,\mathcal{G}) \subseteq L^2(\Omega, \mathcal{G}, \mathbb{P})$).

In general, $\mathbb{E}_{\mathcal{G}}[X] = L^1\text{-}\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[X_n]$ for any $X_n \to X$ in L^1

$$\therefore \quad \begin{aligned} \mathbb{E}_{\mathcal{G}}[X_n]\,Y &\to \mathbb{E}_{\mathcal{G}}[X]\,Y \\ X_n Y &\to XY \end{aligned} \qquad \text{in } L^1$$

$$\therefore \quad \mathbb{E}\big[\mathbb{E}_{\mathcal{G}}[X]\,Y - XY\big] = \lim_{n\to\infty} \mathbb{E}\big[\mathbb{E}_{\mathcal{G}}[X_n]\,Y - X_n Y\big] = 0 .$$

Conversely, if $Z_1, Z_2 \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ each satisfy

$$\mathbb{E}[Z_1 Y] = \mathbb{E}[Z_2 Y] = \mathbb{E}[XY] \qquad \forall\, Y \in B(\Omega, \mathcal{G})$$

then $\mathbb{E}[(Z_1 - Z_2)Y] = 0$. Take $Y = \operatorname{sgn}(Z_1 - Z_2)\,\mathbb{1}_{|Z_1 - Z_2| \le n} \in B(\Omega, \mathcal{G})$

$$\therefore \; 0 = \mathbb{E}\big[|Z_1 - Z_2|\,\mathbb{1}_{|Z_1 - Z_2| \le n}\big] \xrightarrow[\text{DCT}]{n\to\infty} \mathbb{E}\big[|Z_1 - Z_2|\big] \Rightarrow Z_1 = Z_2 \in L^1 .$$

///

Theorem: (Main Properties of Conditional Expectation)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\mathcal{G} \leq \mathcal{F}$ a sub-σ-field. The linear transformation

$$\mathbb{E}_{\mathcal{G}} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{G}, \mathbb{P})$$

satisfies:

✓ 1. (Monotonicity) if $X \leq Y$ a.s. $[\mathbb{P}]$ then $\mathbb{E}_{\mathcal{G}}[X] \leq \mathbb{E}_{\mathcal{G}}[Y]$ a.s. $[\mathbb{P}]$. $\mathbb{E}_{\mathcal{G}}[M] = M$

2. (Δ-ineq.) $|\mathbb{E}_{\mathcal{G}}[X]| \leq \mathbb{E}_{\mathcal{G}}[|X|]$ a.s. $[\mathbb{P}]$. If $|X| \leq M$ a.s. $|\mathbb{E}_{\mathcal{G}}[X]| \leq M$ a.s.

✓ 3. (Averaging) $\mathbb{E}[\mathbb{E}_{\mathcal{G}}[X] Y] = \mathbb{E}[XY]$ $\forall Y \in \mathcal{B}(\Omega, \mathcal{G})$

4. (Product Rule) If $Y \in \mathcal{B}(\Omega, \mathcal{G})$, $\mathbb{E}_{\mathcal{G}}[XY] = \mathbb{E}_{\mathcal{G}}[X] \cdot Y$.

5. (Tower Property) If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ are all σ-fields, then

$$\mathbb{E}_{\mathcal{G}}[\mathbb{E}_{\mathcal{H}}[X]] = \mathbb{E}_{\mathcal{H}}[\mathbb{E}_{\mathcal{G}}[X]] = \mathbb{E}_{\mathcal{H}}[X].$$

$Z = \mathbb{E}_{\mathcal{G}}[Y - X]$

$\downarrow$

Pf. 1. $\mathbb{E}[\mathbb{E}_{\mathcal{G}}[X] \mathbb{1}_B] = \mathbb{E}[X \mathbb{1}_B] \leq \mathbb{E}[Y \mathbb{1}_B] = \mathbb{E}[\mathbb{E}_{\mathcal{G}}[Y] \mathbb{1}_B]$ $\forall B \in \mathcal{G}$.

So, suffices to show that, if $Z \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ and $\mathbb{E}[Z; B] \geq 0$ $\forall B \in \mathcal{G}$ then $Z \geq 0$ a.s. $[\mathbb{P}]$

$B = \{Z < 0\} \in \mathcal{G}$ $\therefore 0 \leq \mathbb{E}[Z \mathbb{1}_{\{Z<0\}}] \leq 0 \Rightarrow \mathbb{E}[Z \mathbb{1}_{\{Z<0\}}] = 0$. $\nearrow$ $\mathbb{P}(Z<0) = 0$.

$0 = \mathbb{E}[-Z \mathbb{1}_{\{Z<0\}}] = \mathbb{E}[|Z| \mathbb{1}_{\{Z<0\}}] \Rightarrow |Z| \mathbb{1}_{\{Z<0\}} = 0$ a.s.,

2. $X \leq |X|$ and $-X \leq |X|$ $\therefore$ by (1) $\downarrow$

$\therefore$ $\mathbb{E}_{\mathcal{G}}[X] \leq \mathbb{E}_{\mathcal{G}}[|X|]$ and $\mathbb{E}_{\mathcal{G}}[-X] \leq \mathbb{E}_{\mathcal{G}}[|X|]$ a.s.

$$\Downarrow \quad |\mathbb{E}_{\mathcal{G}}[X]| \leq \mathbb{E}_{\mathcal{G}}[|X|] \text{ a.s.}$$

3. ✓

4. Let $Y, Z \in B(\Omega, \mathcal{G})$. Then

$$\mathbb{E}[\mathbb{E}_{\mathcal{G}}[X](Y \cdot Z)] = \mathbb{E}[\underline{\mathbb{E}_{\mathcal{G}}[X]} \cdot \underline{YZ}] \overset{(3)}{=} \mathbb{E}[\underline{X}Y\underline{Z}] \overset{(3)}{=} \mathbb{E}[\mathbb{E}_{\mathcal{G}}[XY]Z]$$

$$\therefore \quad \mathbb{E}[(\underbrace{Y\mathbb{E}_{\mathcal{G}}[X] - \mathbb{E}_{\mathcal{G}}[XY]}_{u})Z] = 0 \quad \forall Z \in B(\Omega, \mathcal{G})$$

Take $Z = \operatorname{sgn} u \, \mathbb{1}_{|u| \leq n}$

$$\Rightarrow \quad 0 = \mathbb{E}[|u| \mathbb{1}_{|u| \leq n}] \xrightarrow{DCT} \mathbb{E}[|u|]$$

$$\Rightarrow u = 0 \text{ a.s.}$$

5. $\mathbb{E}_{\mathcal{G}}[\mathbb{E}_{\mathcal{H}}[X]] = \mathbb{E}_{\mathcal{H}}[\mathbb{E}_{\mathcal{G}}[X]] = \mathbb{E}_{\mathcal{H}}[X]$ holds for $X \in L^2$ by orth. proj. thm. Now approximate $X \in L^1$ by $X_n \in L^2$, and be careful.