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 $\|P_K(X)\| \leq \|X\|$, 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. $X - Z \perp K$

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

if $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$

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:

Closed:

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

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

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

If X is $\mathcal{G}$-measurable, then $X^{-1}\{t\} \in$

for each $t \in \mathbb{R}$. In particular, with $t = X(1)$

In fact, for any $\mathbb{P}$, $L^2(\Omega, \mathcal{F}, \mathbb{P}) = \mathbb{R}^{\{1,2,3\}}$

$$L^2(\Omega, \mathcal{G}, \mathbb{P})$$

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

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 $Z = \operatorname{sgn} Y = \begin{cases} 1, & Y > 0 \\ 0, & Y = 0 \\ -1, & Y < 0 \end{cases}$

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

$$\mathbb{E}[|Y|] = \mathbb{E}[Y \cdot \operatorname{sgn} Y]$$

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

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

$\mathbb{E}_{\mathcal{G}}$ 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 $\therefore$ in L^2, and $\|X - X\mathbb{1}_{|X| \leq n}\|_{L^1} = \mathbb{E}[|X| \mathbb{1}_{|X| > n}] \underset{n\to\infty}{\to} 0$

Def: If $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$, define $\mathbb{E}_{\mathcal{G}}[X]$ 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 $\mathbb{E}_{\mathcal{G}}[X] := L^1\text{-}\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[X_n] = L^1\text{-}\lim_{n\to\infty} P_{L^2(\Omega, \mathcal{G}, \mathbb{P})}(X_n)$

- Exists: $\|\mathbb{E}_{\mathcal{G}}[X_n] - \mathbb{E}_{\mathcal{G}}[X_m]\|_{L^1}$

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

$\|\mathbb{E}_{\mathcal{G}}[X_n] - \mathbb{E}_{\mathcal{G}}[Y_n]\|_{L^1}$

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:

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

Pf. If $X \in L^2$, $(\star)$ holds by def^n 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

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

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

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 $\mathbb{E}[(Z_1 - Z_2)Y] = 0.$

Theorem: (Main Properties of Conditional Expectation)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\mathcal{G} \subseteq \mathcal{F}$ 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. $[\mathbb{P}]$.
2. (Δ-ineq.) $|\mathbb{E}_{\mathcal{G}}[X]| \leq \mathbb{E}_{\mathcal{G}}[|X|]$ a.s. $[\mathbb{P}]$.
3. (Averaging) $\mathbb{E}[\mathbb{E}_{\mathcal{G}}[X] Y] = \mathbb{E}[XY]$ $\forall Y \in \mathbb{B}(\Omega, \mathcal{G})$
4. (Product Rule) If $Y \in \mathbb{B}(\Omega, \mathcal{G})$, $\mathbb{E}_{\mathcal{G}}[XY] = \mathbb{E}_{\mathcal{G}}[X] \cdot Y$.
5. (Tower Property) If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ are all σ-fields, then

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

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. $[\mathbb{P}]$

2. $X \le |X|$ and $-X \le |X|$

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

3. ✓

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

$$\mathbb{E}[\mathbb{E}_{\mathcal{G}}[X] Y \cdot Z] = \mathbb{E}[\mathbb{E}_{\mathcal{G}}[X] \cdot YZ] = \mathbb{E}[XYZ]$$

$$\therefore \mathbb{E}[(Y \mathbb{E}_{\mathcal{G}}[X] - \mathbb{E}_{\mathcal{G}}[XY]) Z]$$

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.