Probability Theory

VIII Conditional Expectation

Seongho, Joo

August 7, 2025

Seoul National University

Orthogonal Projection

If H is a finite dimensional inner product space, and $K \subseteq H$ is any subspace, there is an orthogonal projection $P_k : H \to K$ with the properties that

- $P_k(v) = v \quad \forall v \in K$
- $P_k(w) = 0$ if $w \in K^\top$

If we can find an orthonormal basis $\{e_n\}$ for K, then

$$P_k(v) = \sum_n \langle v, e_n \rangle \, e_n$$

We will use this same idea in the Hilbert space $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$

Hilbert Spaces

A Hilbert Space is a complete inner product space.
E.g. $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$ with $\langle X, Y \rangle := \mathbb{E}\, XY$.
In any inner product space, we have Pythagoras's Thm:

$$\text{If } X \perp Y,\ \|X+Y\|^2 = \|X\|^2 + \|Y\|^2$$

Also, we have the Parallelogram Law:

$$\|X+Y\|^2 + \|X-Y\|^2 = 2(\|X\|^2 + \|Y\|^2)$$

If $K \subset H$ is a linear subspace, it is also an inner product space. K is a Hilbert space if and only if K is closed in H.

Hilbert Spaces

A Hilbert Space is a complete inner product space.
E.g. $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$ with $\langle X, Y \rangle := \mathbb{E}\, XY$.
In any inner product space, we have Pythagoras's Thm:

$$\text{If } X \perp Y,\ \|X+Y\|^2 = \|X\|^2 + \|Y\|^2$$

Also, we have the Parallelogram Law:

$$\|X+Y\|^2 + \|X-Y\|^2 = 2(\|X\|^2 + \|Y\|^2)$$

If $K \subset H$ is a linear subspace, it is also an inner product space. K is a Hilbert space if and only if K is closed in H.

Proposition 1

If $K \subseteq H$ is a closed subspace, and $X \in H$, there is a unique closest element $Y \in K$ to Y.

$$||X - Y||^2 = d(x, K)^2 := \inf_{Z \in K} ||X - Z||^2$$

Proof.

Hilbert Spaces

Proposition 2

The unique closest point $Y \in K$ to X is also the unique element $Y \in K$ satisfying

$$X - Y \perp K \quad \text{i.e. } \langle X - Y, Z \rangle = 0 \; \forall Z \in K$$

Proof. If Y is the closest point,for any $Z \in K$, consider

$$t \mapsto \alpha(t) = \|X - (Y + tZ)\|^2 = \|X - Y\|^2 - 2t \langle X - Y, Z \rangle + \|Z\|^2 t^2$$

by assumption, $\alpha(0) = \min \alpha \implies \alpha'(0) = 0 \implies -2 \langle X - Y, Z \rangle = 0$
Conversely, if $Y \in K$ with $X - Y \perp K$, then for any $Z \in K$,

$$\|X - Z\|^2 = \left\| X - Y + \underbrace{Y - Z}_{\in K} \right\|^2 = \|X - Y\|^2 + \|Y - Z\|^2 \geq \|X - Y\|^2$$

$\implies Y =$ unique minimizer of $d(x, K)$.

Hilbert Spaces

Theorem 1 (Orthogonal Projection)

Given a Hilbert space H and a closed subspace $K \subseteq H$, there is a unique linear transformation $P_k : H \to K$ such that

- P_k is Lip_1-cotinuous
- $P_k(Y) = Y \quad \forall Y \in K$
- $P_k(Z) = 0 \quad \forall Z \in K^{\perp}$
- $\langle P_K(X), Y \rangle = \langle X, P_K(Y) \rangle \quad \forall X, Y \in H$

Moreover, if $L \subseteq K$ is another closed subapce, then $P_K P_L = P_L P_K = P_L$. The transformation P_K, the orthogonal projection onto K, can be defined by $P_K(X)=$ the unique element in K closest to X.

Conditional Expectation

Proposition 3

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

Proof. Closed: If X_n is $\mathcal{G}$-measurable, $X_n \to X$ in L^2, $\exists$ subseq $X_{n_k} \to X$ a.s. $\implies X$ is $\mathcal{G}$-measurable. $\therefore X \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Conditional Expectation

Proposition 3

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

Proof. Closed: If X_n is $\mathcal{G}$-measurable, $X_n \to X$ in L^2, $\exists$ subseq $X_{n_k} \to X$ a.s. $\implies X$ is $\mathcal{G}$-measurable. $\therefore X \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Definition 1

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

Conditional Expectation

What is $\mathbb{E}\, X|\mathcal{G}$? It is the $\mathcal{G}$-measurable r.v. that is <u>closest</u> to X

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

I.e. it is the best guess for X, using only the information in G.
Question:: does it only make sense for L^2?

Conditional Expectation

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

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

I.e. it is the best guess for X, using only the information in G.

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

Lemma 1

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

Proof. Use that $Z = \text{sgn} Y$, where $Y = \mathbb{E}_{\mathcal{G}}\,[X]$ is $\mathcal{G}$-measurable.

Note that $L^2 \subseteq L^1$ is dense; given $X \in L^1$, $X\mathbf{1}_{|X|\leq n}$ is bounded and in L^2, and $\left\|X - X\mathbf{1}_{|X|\geq n}\right\|_{L^1} = \mathbb{E}\,|X|\mathbf{1}_{|X|>n} \to 0$ by DCT.

Extension to L^1 r.v.

Definition 2

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 $\mathbb{E}_{\mathcal{G}}[X] := L^1 \ \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[X_n]$

- Note that $\{\mathbb{E}_{\mathcal{G}}[X_n]\}_{n=1}^{\infty}$ is L^1-Cauchy.
- . if $X_n, Y_n \to X$ in L^1, then

$$\|\mathbb{E}_{\mathcal{G}}[X_n] - \mathbb{E}_{\mathcal{G}}[Y_n]\|_{L^1} \leq \|X_n - Y_n\|_{L^1} \to 0 \text{ as } n \to \infty$$

therefore, it is well-defined.

Averaging property

Proposition 4

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:

$$\mathbb{E}\,\mathbb{E}_{\mathcal{G}}[X]\,Y = \mathbb{E}\,XY \quad \forall Y \in \mathbb{B}(\Omega, \mathcal{G}) \cdots (\dagger)$$

Proof. For the converse, take $Y = \mathrm{sgn}(Z_1 - Z_2)\mathbf{1}_{|Z_1 - Z_2| \leq n} \in \mathbb{B}(\Omega, \mathcal{G})$

Averaging property

Proposition 4

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:

$$\mathbb{E}\, \mathbb{E}_{\mathcal{G}}[X]\, Y = \mathbb{E}\, XY \quad \forall Y \in \mathbb{B}(\Omega, \mathcal{G}) \cdots (\dagger)$$

Proof. For the converse, take $Y = \mathrm{sgn}(Z_1 - Z_2)\mathbf{1}_{|Z_1 - Z_2| \le n} \in \mathbb{B}(\Omega, \mathcal{G})$

Theorem 2 (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 $\mathbb{E}_{\mathcal{G}}$ satisfies:

1. (Monotonicity) if $X \le Y$ a.s. then $\mathbb{E}_{\mathcal{G}}[X] \le \mathbb{E}_{\mathcal{G}}[Y]$ a.s.
2. ($\triangle$ -ineq) $|\mathbb{E}_{\mathcal{G}}[X]| \le \mathbb{E}_{\mathcal{G}}[|X|]$
3. (Averaging) $\mathbb{E}\, \mathbb{E}_{\mathcal{G}}[X]\, Y = \mathbb{E}\, XY \quad \forall Y \in \mathbb{B}(\Omega, \mathbb{G})$
4. (Product Rule) If $Y \in \mathbb{B}(\Omega, \mathcal{G})$, $\mathbb{E}_{\mathcal{G}}[XY] = \mathbb{E}_{\mathcal{G}}[X]\, Y$
5. (Tower Property) If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ are all σ-fields, then
$$\mathbb{E}_{\mathcal{G}}[\mathbb{E}_{\mathcal{G}}[X]] = \mathbb{E}_{\mathcal{H}}[\mathbb{E}_{\mathcal{G}}[X]] = \mathbb{E}_{\mathcal{H}}[X]$$

Averaging property

Lemma 2

$Z = \mathbb{E}_{\mathcal{G}}[X]$ if and only if $Z \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ and $\mathbb{E}\, Z : B = \mathbb{E}\, X : B \ \forall B \in \mathcal{G}$

Proof. For the converse, use the Dynkin's Multiplicative System theorem.

Conditional Expectation

Theorem 3

$\mathbb{E}_{\mathcal{G}}$ satisfies the standard integral convergence results:

- (cMCT) If $0 \leq X_n \leq X_{n+1}$, then $\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[X_n] = \mathbb{E}_{\mathcal{G}}[\lim_{n\to\infty} X_n]$ a.s.
- (CFatou) If $X_n \geq 0$ a.s, then $\mathbb{E}_{\mathcal{G}}[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}[X_n]$ a.s
- (cDCT) If $X_n \to X$ a.s. and $|X_n| \leq Y \in L^1$ a.s., then $\mathbb{E}_{\mathcal{G}}[X_n] \to \mathbb{E}_{\mathcal{G}}[X]$ a.s. and L^1.

Theorem 4 (Conditional Jensen's Inequality)

Let $X_1 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G} \subseteq \mathcal{F}$ is sub-σ-field. If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex, and $\varphi(X) \in L^1$, then
$\varphi(\mathbb{E}_{\mathcal{G}}[X]) \in L^1(\Omega, \mathcal{G}, \mathbb{P})$, and $\varphi(\mathbb{E}_{\mathcal{G}}[X]) \leq \mathbb{E}_{\mathcal{G}}[\varphi(X)]$ a.s.

Corollary 1

For $1 \leq p < \infty, \mathbb{E}_{\mathcal{G}} : L^p \to L^p$ is a contraction.

Proof. Given convex $\varphi_p(x) = |x|^p$, $\varphi_p(\mathbb{E}_{\mathcal{G}}[X]) \leq \mathbb{E}_{\mathcal{G}}[\varphi_p(X)]$
$\implies \|\mathbb{E}_{\mathcal{G}}[X]\|_p \leq ||X||_p$

Conditional Expectation and Independence

$\mathbb{E}_{\mathcal{G}}[X]$ is the best guess at X using only information in $\mathcal{G}$. What if $\mathcal{G}$ has no information about X?

Proposition 5

Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{B})$ be a random variable, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub σ-field. If $\sigma(X), \mathcal{G}$ are independent, and $f : S \to \mathbb{R}$ is such that $f(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then

$$\mathbb{E}_{\mathcal{G}}[f(X)] = \mathbb{E}\, f(X) \text{ a.s. } \cdots (\dagger)$$

Conversely. if $\dagger$ holds for all $f \in \mathbb{B}(S, \mathcal{B})$, then $\sigma(X), \mathcal{G}$ are independent.

Conditional expectation and independence

If X, Y are independent, Y is constant w.r.t X. We can make this precise as follows:

Proposition 6

Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{B}). Y : (\Omega, \mathcal{F}) \to T, \mathcal{C})$ be random variables. let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$. If X, Y are independent, and $f \in \mathbb{B}\,(S \times T, \mathcal{B} \otimes \mathcal{C})$, then

$$\mathbb{E}\, f(X, Y) | X = x = \mathbb{E}\, f(x, Y)$$

$$\mathbb{E}\, f(X, Y) | X = \mathbb{E}\, f(x, Y)|_{x=X}$$

Question: What form the conditional expectation has w.r.t. the joint density $\rho(x, y)$?

Conditional expectation and density

Proposition 7

Let (X, Y) have density $\rho = \rho_{X,Y}$. Let $\rho_X(x) = \int \rho_{X,Y}(x, y)\,\mathrm{d}y$ be the marginal density of X. Define

$$\rho_{Y|X}(y|x) := \mathbf{1}_{0<\rho_X<\infty} \frac{\rho_{X,Y}(x, y)}{\rho_X(x)}$$

Then for $f \in \mathbb{B}(\mathbb{R}^2)$,

$$\mathbb{E}\, f(X, Y)|X = g(X), \quad g(x) = \int f(x, y)\rho_{Y|X}(y|x)\,\mathrm{d}y$$

Proof. Note that we see that g satisfies

$$g(x)\rho_X(x) = \int f(x, y)\rho_{X,Y}(x, y)\,\mathrm{d}y \quad \text{for a.e. } x \in \mathbb{R}$$