Lecture 18: L_p spaces of r.v.s

Just as in the previous lecture we will be fixing a probability space $(\Omega, \mathcal{F}, P)$ and consider the collection of r.v.s defined on that space.

In particular:

Assumption for the remainder of this lecture:

Suppose $X, Y, X_1, X_2, \dots$ are r.v.s all defined on the same probability space

$$(\Omega, \mathcal{F}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$
$$(\Omega, \mathcal{F}, P) \xrightarrow{Y} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$
$$\vdots$$

L_p spaces $(p \geq 1)$

Definition: Let $(\Omega, \mathcal{F}, P)$ be a probability space and $p \geq 1$.
The L_p space of r.v.s defined on $(\Omega, \mathcal{F}, P)$ is defined as

$$\{X : \Omega \to \mathbb{R} \text{ s.t. } X \text{ (m) } \mathcal{F}/\mathcal{B}(\mathbb{R}) \ \& \ E|X|^p < \infty\}$$

and denoted $L_p(\Omega, \mathcal{F}, P) \equiv L_p(P) \equiv L_p$.

Remark: We work with random variables but most of the following results can be extended to the set of random vectors all mapping into $\mathbb{R}^d$.

We will be interested in the metric & geometric properties of L_p & interpreting some classic functional analysis results from a probabilistic perspective.

Example:

Let W_t be Brownian Motion so that

$$(\Omega, \mathcal{F}, P) \xrightarrow{W \text{ (m)}} (C[0,\infty), \mathcal{B}(C[0,\infty)))$$

Since for each fixed $t \in [0,\infty)$ W_t is a r.v. defined on $(\Omega, \mathcal{F}, P)$ we can consider the stochastic process $(W_t : t \in [0,\infty))$ as a collection of r.v.s indexed by t

$$\{W_t : t \in [0,\infty)\} \subset L_2(\Omega, \mathcal{F}, P)$$

Definition:

For $X, Y \in L_p(\Omega, \mathcal{F}, P)$ define

$$\|X\|_p := (E|X|^p)^{1/p}$$
$$d_p(X,Y) := \|X - Y\|_p$$

Theorem (Hölder)

For any two r.v.s X & Y defined on $(\Omega, \mathcal{F}, P)$ and $p, q > 1$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$E|XY| \leq \underbrace{\|X\|_p \|Y\|_q}_{\text{could be } 0 \cdot \infty = 0 \text{ by convention}}$$

Moreover if $X, Y \in \mathcal{L}(\Omega, \mathcal{F}, P)$ then

$$|E(XY)| \leq \|X\|_p \|Y\|_q$$

Proof: We already proved this in Lecture 11.

Theorem:

$$1 \le p < q \implies L_q(\Omega, \mathcal{F}, P) \subset L_p(\Omega, \mathcal{F}, P).$$

Proof: Like Hölder, this comes from Young's inequality: $a^{w_1} b^{w_2} \le w_1 a + w_2 b$ when $w_1, w_2, a, b > 0$ and $w_1 + w_2 = 1$.

Indeed set $w_1 = \frac{p}{q} < 1$ & $w_2 = 1 - w_1$. Then

$$X \in L_q \implies E|X|^q = E|X|^{p\frac{q}{p}}$$
$$= E\left(\left(|X|^p\right)^{w_1} 1^{w_2}\right)$$
$$\le \underbrace{w_1 E|X|^p + w_2}_{<\infty}$$

QED.

Theorem: ($\|\cdot\|_p$ is a pseudo-norm)

$\forall X \in L_p(\Omega, \mathcal{F}, P)$ we have that

(i) $\|X\|_p \ge 0$

(ii) $\|X\|_p = 0 \implies X = 0$ P-a.e. hence its only pseudo. By "a.e. useful" Thm.

(iii) $\|cX\|_p = |c|\,\|X\|_p \quad \forall c \in \mathbb{R}$

(iv) $\|X+Y\|_p \le \|X\|_p + \|Y\|_p$ (Minkowski's ineq)

Proof:

we just need to show (iv).

$$E|X+Y|^p = E\left(|X+Y|\,|X+Y|^{p-1}\right)$$
$$\le E\left(|X|\,|X+Y|^{p-1}\right) + E\left(|Y|\,|X+Y|^{p-1}\right)$$

Now notice $\frac{1}{p} + \frac{1}{q} = 1 \implies 1 + \frac{p}{q} = p$
$\implies \frac{p}{q} = p - 1$

$$= E\left(|X|\,|X+Y|^{p/q}\right) + E\left(|X|\,|X+Y|^{p/q}\right)$$
$$\overset{\text{Hölder}}{\le} \|X\|_p \left\| |X+Y|^{p/q} \right\|_q + \|Y\|_p \left\| |X+Y|^{p/q} \right\|_q$$
$$= \left(\|X\|_p + \|Y\|_p\right) \underbrace{\left(E|X+Y|^p\right)^{1/q}}_{\text{divide this out of both sides}}$$

$$\therefore \underbrace{\left(E|X+Y|^p\right)^{1-\frac{1}{q}}}_{=\|X+Y\|_p} \le \|X\|_p + \|Y\|_p$$

QED.

Remark: The previous Thm shows $d_p(X,Y)$ is a pseudo-metric on L_p.

It will also be useful to note that $\|\cdot\|_p$ is continuous w.r.t d_p.

Theorem: $X, Y \in L_p \implies \left| \|X\|_p - \|Y\|_p \right| \le d_p(X,Y)$

Proof:

Minkowski: $\|X\|_p \le \|X-Y\|_p + \|Y\|_p$

$$\|Y\|_p \le \underbrace{\|X-Y\|_p}_{=d_p(X,Y)} + \|X\|_p$$

QED

L_p convergence

Here we study completeness, closure & separability of L_p and prove the "L_p convergence theorem" which will be useful later.

Definition:

$X_n \xrightarrow{L_p} X$ iff $E|\underbrace{X_n - X}|^p \to 0$ as $n \to \infty$.

technically no requirement that $X_n, X \in L_p$

Theorem (uniqueness of limits)

$$X_n \xrightarrow{L_p} X \ \& \ X_n \xrightarrow{L_p} Y \Rightarrow X = Y \ P\text{-a.e.}$$

Proof:

Note the following useful identity which follows by convexity of $|\cdot|^p$

$$\left|\tfrac{x+y}{2}\right|^p \le \tfrac{1}{2}|x|^p + \tfrac{1}{2}|y|^p$$

$$\therefore \quad E|\underset{\pm X_n}{X-Y}|^p \le 2^p\Big(\underbrace{\tfrac{1}{2}E|X-X_n|^p + \tfrac{1}{2}E|Y-X_n|^p}_{\to 0 \text{ as } n\to\infty}\Big) \quad QED$$

Theorem: (Cauchy Criteria)

$X_n \xrightarrow{L_p}$ to some r.v. X iff

$$\lim_n \lim_m E|X_m - X_n|^p = 0$$

Proof:

$(\Longrightarrow)$

$$E|\underset{\pm X}{X_m - X_n}|^p \le 2^p\Big(\underbrace{E\left|\tfrac{X_m-X}{2}\right|^p + E\left|\tfrac{X_n-X}{2}\right|^p}_{\to 0 \text{ as } m,n\to\infty}\Big)$$

$(\Longleftarrow)$

$$P(|X_m - X_n| \ge \varepsilon) \le \frac{E|X_m-X_n|^p}{\varepsilon^p}$$

implies $\{X_n\}_{n\ge1}$ is Cauchy for convergence in probability.

$\therefore\ \exists$ r.v. X s.t. $X_n \xrightarrow{P} X$

$\therefore\ \exists n_k$ s.t. $X_{n_k} \xrightarrow[k\to\infty]{a.e.} X$ by sub-sub-seq Thm.

$$\therefore \quad |X_n - X_{n_k}|^p \xrightarrow[k\to\infty]{a.e.} |X_n - X|^p \quad \forall n$$

by continuous mapping since $X_n - X_{n_k} \xrightarrow{a.e.} X_n - X$

Now

$$\begin{aligned} E|X_n - X|^p &\le \liminf_k E|X_n - X_{n_k}|^p, \text{ Fatou} \\ &\le \limsup_k E|X_n - X_{n_k}|^p \\ &\le \limsup_m E|X_n - X_m|^p \end{aligned}$$

Taking $\lim_n$ of both sides gives

$$X_n \xrightarrow{L_p} X. \qquad QED$$

Theorem (L_p is Polish w.r.t d_p)

If $p \ge 1$ then $L_p(\Omega, \mathcal{F}, P)$ is a linear space which is closed & complete w.r.t d_p.

If, in addition, $\mathcal{F}$ is countably generated then $L_p(\Omega,\mathcal{F},P)$ is separable.

Proof:

(L_p is linear): Follows by $|X+Y|^p \le 2^p(\frac{1}{2}|X|^p + \frac{1}{2}|Y|^p)$

(L_p is closed): if $X_n \in L_p$ & $X_n \xrightarrow{L_p} X$ then

$$|X|^p \le 2^p\left(\left|\tfrac{X_n}{2}\right|^p + \left|\tfrac{X-X_n}{2}\right|^p\right)$$

Taking expected value of both sides gives the result.

(L_p is complete): Follows by the Cauchy Criteria thm.

(L_p is separable):

Suppose $\mathcal{F} = \sigma\langle \mathcal{A} \rangle$ where $\mathcal{A}$ is a countable collection of generators.

Let $X \in L_p$. By the structure thm of Lecture 9 $\exists$ bdd simple X_n's s.t.

$$X_n \xrightarrow{a.e.} X$$

where $X_n \in L_p$ by bddness.

Also, although not explicitly stated in the structure Thm, the X_n's satisfy $|X_n| \le |X|$

$$\therefore \; |X - X_n|^p \le 2^p\left(\frac{|X|^p}{2} + \frac{|X_n|^p}{2}\right)$$

$$\le 2^p |X|^p$$

so by the DCT we have

$$E|X_n - X|^p \to 0 \text{ as } n \to \infty.$$

Since X_n is simple, it has the form

$$X_n = \sum_{k=1}^{m} c_k I_{F_k}, \quad F_k \in \mathcal{F} = \sigma\langle \mathcal{Q} \rangle$$

using a result in Lecture 5 we can find $\hat{F}_k \in \mathcal{Q}$, $\hat{c}_k \in \mathbb{Q}$ s.t.

$$\| X_n - \hat{X}_n \|_p \le \frac{1}{n}$$

where $\hat{X}_n = \sum_{k=1}^{m} \hat{c}_k I_{\hat{F}_k}$ (hint: choose $\hat{F}_k$ so that $P(F_k \triangle \hat{F}_k) < \left[\frac{(1/n)}{2^k |c_k|}\right]^p$).

For this $\hat{X}_n$ we have

$$\| X - \hat{X}_n \|_p \to 0 \;\; \& \;\; \hat{X}_n \in L$$

QED

Recall the definition of uniform integrability (UI) specialized to r.v.s:

$X_1, X_2, \dots$ are UI iff

$$\lim_{c \to \infty} \sup_n E\left(|X_n| I_{|X_n| \ge c}\right) = 0$$

when talking about limits its understood we can drop any finite number of X_n's

Lets also recall the UI theorems we did in Lecture 10 but specialized to r.v.s

Theorem: (UI for $\lim E = E \lim$)

If $X_n \xrightarrow{a.e.} X$ & the X_n's are UI then $EX_n \to EX$ & $X, X_n \in L_1$

Theorem: (UI converse)

If $X_n \xrightarrow{a.e.} X$ & $EX_n \to EX$ & $X, X_n \in L_1$ then the X_n's are UI.

Here is our L_p convergence Thm which effectively shows

$$\xrightarrow{L_p} \; \equiv \; \xrightarrow{P} + |X_n|^p\text{'s are UI}$$

Theorem: (L_p convergence Thm)

Let $X_n \in L_p$ for all n. Then the following are equivalent:

(i) $X_n \xrightarrow{L_p} X$

(ii) $X_n \xrightarrow{P} X$ and $E|X_n|^p \to E|X|^p < \infty$

(iii) $X_n \xrightarrow{P} X$ and the $|X_n|^p$'s are UI

Proof:

(i) $\Rightarrow$ (ii)

we already know $X_n \xrightarrow{P} X$ by Markov's ineq.
$X \in L_p$ since L_p is closed. Finally
by $\big| \|X\|_p - \|X_n\|_p \big| \le d_p(X, X_n) \to 0$
we have
$$E|X_n|^p \to E|X|^p < \infty$$

(ii) $\Rightarrow$ (i).

Here is where we use the Probability Sandwich result proved in the last lecture.

$$0 \le |X_n - X|^p \le 2^p \underbrace{\left(\tfrac{1}{2}|X_n|^p + \tfrac{1}{2}|X|^p\right)}_{\downarrow P \;\; 2^p|X|^p \;=: Y} =: Y_n$$

$|X_n - X|^p \xrightarrow{P} 0$ by continuous mapping since $X_n - X \xrightarrow{P} 0$

since $Y_n, Y \in L_1$ & $EY_n \to EY$ by assumption sandwich says that $E|X_n - X|^p \to 0$.

(ii) $\Rightarrow$ (iii):

using the sub-sub-seq characterization of $\xrightarrow{P}$ one can extend the UI converse to require $\xrightarrow{P}$ instead of $\xrightarrow{a.e.}$.

$\therefore$ From (ii) we have $X_n \xrightarrow{P} X$ & by continuous mapping $|X_n|^p \xrightarrow{P} |X|^p$

Also by assump $E|X_n|^p \to E|X|^p < \infty$ so that $X_n, X \in L_p$ for suff large n

$\therefore$ The X_n's are UI by UI converse

(iii) $\Rightarrow$ (ii):

This one similarly follows from an $\xrightarrow{P}$ version of the UI theorem

QED.

Hilbert Space Geometry of $L_2(\Omega, \mathcal{F}, P)$

For L_2 Hölder gives $E(XY) \le \|X\|_2 \|Y\|_2 < \infty$

$\therefore$ we can define an inner product on L_2 defined as
$$\langle X, Y \rangle := E(XY)$$

Much of what statisticians do basically corresponds to geometric operations w.r.t. $\langle \cdot, \cdot \rangle$.
The geometry of $(L_2, \langle \cdot, \cdot \rangle)$ is the geometry of variation & co-variation:
i.e. when $E(X) = E(Y) = 0$ then
$$\langle X, Y \rangle = \operatorname{cov}(X, Y)$$
$$\langle X, X \rangle = \|X\|_2^2 = \operatorname{var}(X)$$
$$\|X\|_2 = sd(X).$$

Basic Properties of $\langle \cdot, \cdot \rangle$:
$\forall X, Y \in L_2(\Omega, \mathcal{F}, P)$

(1) $\langle X, X \rangle \ge 0$

(2) $\langle X, X \rangle > 0$ unless $X = 0$ P-a.e.

(3) $\langle X, Y \rangle = \langle X, Y \rangle$

(4) $\langle X, Y + \alpha Z \rangle = \langle X, Y \rangle + \alpha \langle X, Z \rangle$

(5) $|\langle X, Y \rangle| \le \|X\|_2 \|Y\|_2$

(6) $X_n \xrightarrow{L_2} X \Rightarrow \langle X_n, Y \rangle \to \langle X, Y \rangle$

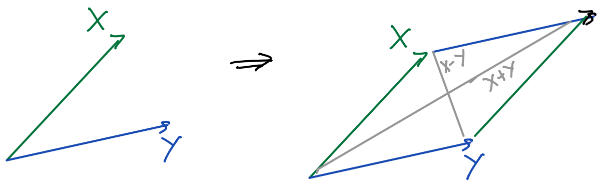
which is true since
$$|\langle X_n, Y \rangle - \langle X, Y \rangle| = |\langle X_n - X, Y \rangle| \le \|X_n - X\|_2 \|Y\|_2$$

(7) $\|X + Y\|_2^2 = \|X\|_2^2 + 2\langle X, Y \rangle + \|Y\|_2^2$

(8) $\|X+Y\|_2^2 + \|X-Y\|_2^2 = 2\|X\|_2^2 + 2\|Y\|_2^2$

This follows by adding

$$\|X+Y\|_2^2 = \|X\|_2^2 + 2\langle X,Y\rangle + \|Y\|_2^2$$
$$\|X-Y\|_2^2 = \|X\|_2^2 - 2\langle X,Y\rangle + \|Y\|_2^2$$



To see an example of the L_2 geometry viewpoint in estimation problems suppose $Z(x)$ is a mean zero continuous Gaussian random field defined on a region $S \subset \mathbb{R}^d$.
i.e. there exists $(\Omega, \mathcal{F}, P)$ s.t.

$$(\Omega, \mathcal{F}, P) \xrightarrow{(Z(x):\, x\in\mathbb{R}^d)} (C(\mathbb{R}^d), \mathcal{B}(C(\mathbb{R}^d)))$$

and $Z(x)$ has Gaussian f.d.d.s & $E(Z(x)) = 0$ $\forall x \in S$

$\therefore$ The collection of r.v.s $Z(x)$ indexed by $x \in S$ satisfies

$$\{Z(x) : x \in S\} \subset L_2(\Omega, \mathcal{F}, P)$$

In random field theory we often study the following Hilbert space

$$\begin{aligned} L(S) &:= \text{closed linear span (in } L_2) \text{ of } \{Z(x) : x \in S\} \\ &= \text{closure with } L_2 \text{ limits of } \Big\{\sum_{k=1}^n c_k Z(x_k) : x_k \in S,\ c_k \in \mathbb{R}\Big\} \\ &\subset L_2(\Omega, \mathcal{F}, P) \end{aligned}$$

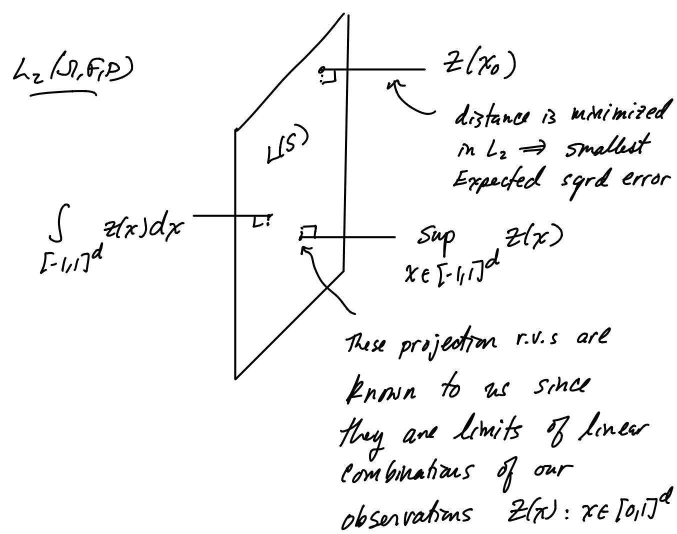
Now "Best linear prediction" (BLP) of some unobserved $L_2(\Omega,\mathcal{F},P)$ r.v. simply given by projection:

e.g. Suppose $Z(x)$ is defined on $x \in \mathbb{R}^d$ but only observed on $x \in [0,1]^d$.

If we want to predict things like

- $Z(x_0)$ for $x_0 \notin [0,1]^d$
- $\int_{[-1,1]^d} Z(x)\,dx$
- $\sup_{x \in [-1,1]^d} Z(x)$

So long as these r.v.s are in L_2 the BLP is a projection $L(S)$



Definition $X \in L_2$ is orthogonal to $Y \in L_2$ iff $\langle X, Y \rangle = 0$ (denoted by $X \perp Y$).

Theorem: (Projection Thm)

Let S be a closed linear subspace of L_2 & $Y \in L_2$. Then $\exists$ a P-a.e. unique $\mathcal{P}_S Y \in S$ s.t.

$$\|Y - \mathcal{P}_S Y\|_2 = \inf_{X \in S} \|Y - X\|_2.$$

Moreover $\mathcal{P}_S Y$ is characterized by the following two properties

(1) $\mathcal{P}_S Y \in S$

(2) $\underbrace{(Y - \mathcal{P}_S Y)}_{\text{prediction residual.}} \perp X \quad \forall X \in S$

Proof:

(Find $\mathcal{P}_S Y$) Let $X_n \in S$ s.t.

$$\|Y - X_n\|_2 \xrightarrow{n \to \infty} \inf_{X \in S} \|Y - X\|_2$$

Now we show $\{X_n\}_{n \geq 1}$ is Cauchy with the Parallelogram Thm:

$$\|(X_n - Y) + (X_m - Y)\|_2^2 + \|\underbrace{X_n - X_m}_{(X_n - Y) - (X_m - Y)}\|_2^2 = \underbrace{2\|X_n - Y\|_2^2 + 2\|X_m - Y\|_2^2}_{=: I_{nm}}$$

where $\lim_{n m} I_{nm} = 4\left(\inf_{X \in S} \|Y - X\|_2\right)^2$

$$\therefore \|X_n - X_m\|_2^2 = I_{nm} - \|\underbrace{(X_n - Y) + (X_m - Y)}_{2\left(\frac{X_n + X_m}{2} - Y\right)}\|_2^2$$

$$= I_{nm} - 4\Big\|\underbrace{\tfrac{X_n + X_m}{2}}_{\in S \text{ by linearity}} - Y\Big\|_2^2$$

$$\leq I_{nm} - 4\left(\inf_{X \in S} \|X - Y\|_2\right)^2$$

$$\longrightarrow 0 \quad \text{as } n, m \to \infty$$

$\therefore \{X_n\}_{n \geq 1}$ is Cauchy & by completeness

$\exists \mathcal{P}_S Y \in L_2$ s.t. $\underbrace{X_n}_{\in S} \xrightarrow{L_2} \underbrace{\mathcal{P}_S Y}_{\in S}$ implies $\mathcal{P}_S Y \in S$ since S is closed

Also, for this $\mathcal{P}_S Y$ we have

$$\inf_{X \in S} \|X - Y\|_2 \leq \|\mathcal{P}_S Y \underset{\pm X_n}{} - Y\|_2$$

$$\leq \underbrace{\|\mathcal{P}_S Y - X_n\|_2}_{\to 0} + \underbrace{\|X_n - Y\|_2}_{\to \inf_{X \in S} \|X - Y\|_2}$$

$$\therefore \inf_{X \in S} \|X - Y\|_2 = \|\mathcal{P}_S Y - Y\|_2$$

(Show $\mathcal{P}_S Y$ is unique P-a.e.)

Suppose $X_0 \in S$ s.t. $\|X_0 - Y\|_2 = \inf \cdots$

Again by the Parallelogram Thm

$$\|(X_0 - Y) + (\mathcal{P}_S Y - Y)\|_2^2 + \|X_0 \underset{\pm Y}{} - \mathcal{P}_S Y\|_2^2$$

$$= 2\underbrace{\|X_0 - Y\|_2^2}_{2\inf^2} + 2\underbrace{\|\mathcal{P}_S Y - Y\|_2^2}_{= 2\inf^2}$$

$$\therefore \quad \|X_0 - P_S Y\|_2^2 \le 4\inf^2 - \Big\| 2\Big(\underbrace{\frac{X_0 + P_S Y}{2}}_{\in S} - Y\Big)\Big\|_2^2$$

$$\le 4\inf^2 - 4\inf^2 = 0$$

$$\therefore \quad X_0 \overset{P.\,a.e.}{=} P_S Y.$$

(Show $(Y - P_S Y) \perp X, \ \forall X \in S$):

Choose $X \in S$ s.t. $X \neq 0$ a.e. (if $X = 0$ then the result is true).

For $c \in \mathbb{R}$ set

$$f(c) = \|Y - (P_S Y - cX)\|_2^2$$

Let $c_{min} := \underset{c \in \mathbb{R}}{\operatorname{argmin}} f(c)$.

Two ways to compute c_{min}

1st: $c_{min} = 0$ by minimizing properties of $P_S Y$.

2nd:

$$f(c) = \|Y - P_S Y\|_2^2 + 2c\langle Y - P_S Y, X\rangle + c^2\|X\|_2^2$$

$$\therefore \quad f'(c) = 2\langle Y - P_S Y, X\rangle + 2c\|X\|_2^2$$

$$\therefore \quad c_{min} = -\frac{\langle Y - P_S Y, X\rangle}{\|X\|_2^2} \quad \leftarrow \text{need } \|X\|_2^2 > 0$$

$$= 0 \quad \text{since } c_{min} = 0$$

$$\therefore \quad \langle Y - P_S Y, X\rangle = 0$$

(Show $W \in S$ & $(Y - W) \perp X \ \forall X \in S \Rightarrow W = P_S Y$)

$\forall X \in S$ we have

$$\|X - Y\|_2^2 = \|X - W\|_2^2 + \underbrace{2\langle \underbrace{X - W}_{\in S}, W - Y\rangle}_{0} + \|W - Y\|_2^2$$

$\pm W$

$$\therefore \quad \inf_{X \in S} \|X - Y\|_2^2 = \Big[\underbrace{\inf_{X \in S} \|X - W\|_2^2}_{=0 \text{ since } W \in S}\Big] + \|W - Y\|_2^2$$

$$\therefore \quad W = P_S Y$$

QED