

## Lecture 18: $L_p$ spaces of r.v.s

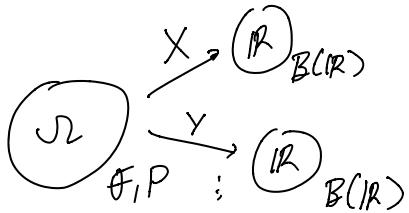
(1)

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



$L_p$  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

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

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

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

$$(\Omega, \mathcal{F}, P) \xrightarrow{W_t} (\mathbb{C}[0, \infty), \mathcal{B}(\mathbb{C}[0, \infty)))$$

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

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

$\underbrace{\quad}_{0 \cdot \infty = 0 \text{ by convention}}$  could be

Moreover if  $X, Y \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$  then

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

Proof: We already proved this in Lecture 11.

Theorem:

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

(3)

Proof: like Hölder, this comes from Young's inequality:  $a^{w_1} b^{w_2} \leq w_1 a + w_2 b$  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

$$\begin{aligned} X \in L_q &\implies E|X|^q = E|X|^{p \cdot \frac{q}{p}} \\ &= E((|X|^p)^{w_1} 1^{w_2}) \\ &\leq w_1 \underbrace{E|X|^p}_{< \infty} + w_2 \end{aligned}$$

QED

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

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

- (i)  $\|X\|_p \geq 0$
- (ii)  $\|X\|_p = 0 \Rightarrow X = 0 \text{ P-a.e.} \leftarrow$  hence its only pseudo.  
By "a.e. useful" norm.
- (iii)  $\|cX\|_p = |c| \|X\|_p \quad \forall c \in \mathbb{R}$
- (iv)  $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p \quad (\text{Minkowski's neg})$

Proof:

We just need to show (iv).

$$\begin{aligned} E|X+Y|^p &= E(|X+Y| |X+Y|^{p-1}) \\ &\leq E(|X| |X+Y|^{p-1}) + E(|Y| |X+Y|^{p-1}) \\ &\quad \text{Now notice } \frac{1}{p} + \frac{1}{q} = 1 \Rightarrow 1 + \frac{p}{q} = p \\ &\quad \Rightarrow \frac{p}{q} = p-1 \\ &= E(|X| |X+Y|^{\frac{p}{q}}) + E(|Y| |X+Y|^{\frac{p}{q}}) \\ &\stackrel{\text{Hölder}}{\leq} \|X\|_p \| |X+Y|^{\frac{p}{q}} \|_q + \|Y\|_p \| |X+Y|^{\frac{p}{q}} \|_q \\ &= (\|X\|_p + \|Y\|_p) \underbrace{\left( E|X+Y|^p \right)^{\frac{1}{q}}}_{\text{divide this out of both sides}} \end{aligned}$$

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

QED.

(4)

Remark: The previous Thm shows that  $d_p(X, Y)$  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:  $X, Y \in L_p \implies |\|X\|_p - \|Y\|_p| \leq d_p(X, Y)$

Proof:

$$\text{Minkowski: } \|X\|_p \leq \|X-Y\|_p + \|Y\|_p$$

$$\|Y\|_p \leq \|X-Y\|_p + \|X\|_p$$

$$\underbrace{= d_p(X, Y)}$$

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  $\underbrace{E|X_n - X|^p}_{\text{technically no requirement}} \rightarrow 0$  as  $n \rightarrow \infty$ .  
that  $X_n, X \in L_p$

### Theorem (uniqueness of limits)

(5)

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

Proof:

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

$$\left| \frac{x+y}{2} \right|^p \leq \frac{1}{2} |x|^p + \frac{1}{2} |y|^p$$

$$\therefore E|X-Y|^p = 2^p \underbrace{\left( \frac{1}{2} E|X-X_n|^p + \frac{1}{2} E|Y-X_n|^p \right)}_{\stackrel{\exists X_n}{\rightarrow 0 \text{ as } n \rightarrow \infty}} \quad \text{QED}$$

### 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:

$$( \Rightarrow ) \quad E|X_m - X_n|^p \leq 2^p \left( \underbrace{E\left(\frac{|X_m - X|}{2}\right)^p}_{\rightarrow 0} + \underbrace{E\left(\frac{|X - X_n|}{2}\right)^p}_{\rightarrow 0 \text{ as } m, n \rightarrow \infty} \right)$$

( $\Leftarrow$ )

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

implies  $\{X_n\}_{n \geq 1}$  is Cauchy for convergence in probability.

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

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

$\therefore |X_n - X_{n_p}|^p \xrightarrow[k \rightarrow \infty]{a.e.} |X_n - X|^p$  & n  
by continuous mapping since  
 $X_n - X_{n_p} \xrightarrow{k \rightarrow \infty} X_n - X$

Now

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

Taking  $\lim_n$  of both sides gives

$$X_n \xrightarrow{L^p} X. \quad \text{QED}$$

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

If  $p \geq 1$  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  $|X+Y|^p \leq 2^p \left( \frac{1}{2} |X|^p + \frac{1}{2} |Y|^p \right)$

( $L_p$  is closed): If  $X_n \in L_p$  &  $X_n \xrightarrow{L^p} X$

$$\text{then } |X|^p \leq 2^p \left( \frac{1}{2} |X_n|^p + \frac{1}{2} |X - X_n|^p \right)$$

Taking expected value of both sides gives the result.

( $L_p$  is complete): Follows by the Cauchy criteria then.

( $L_p$  is separable):

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

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

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

where  $X_n \in L_p$  by boldness.

Also, although not explicitly stated in the (7) structure Thm, the  $X_n$ 's satisfy  $|X_n| \leq |X|$

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

so by the DCT we have

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

Since  $X_n$  is simple, it has the form

$$X_n = \sum_{k=1}^m c_k I_{F_k}, \quad F_k \in \mathcal{F} = \sigma\langle \mathcal{A} \rangle \\ = \delta\langle f(\mathcal{A}) \rangle$$

using a result in Lecture 5 we can find

$$\hat{f}_n \in f(\mathcal{A}), \quad \hat{c}_n \in \mathbb{Q} \text{ s.t.}$$

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

$$\text{where } \hat{X}_n = \sum_{k=1}^m \hat{c}_k I_{\hat{F}_n} \quad (\text{hint: choose}$$

$$\hat{F}_n \text{ so that } P(F_k \Delta \hat{F}_n) < \left[ \frac{\epsilon_n}{2^k |c_k|} \right]^p.$$

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

$$\|X - \hat{X}_n\|_p \rightarrow 0 \text{ & } \hat{X}_n \in L.$$

Since any field generate by a countable collection of events is countable,  $f(\mathcal{A})$  is countable.

$\therefore$  the collection of all such approximating  $\hat{X}_n$ 's forms a dense countable subset of  $L_p(\Omega, \mathcal{F}, P)$ . QED

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

$X_1, X_2, \dots$  are UI iff

$$\lim_{c \rightarrow \infty} \sup_n E(|X_n| I_{|X_n| \geq c}) = 0$$

*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  $EX_n \rightarrow EX$  &  $X, X_n \in L$ ,

**Theorem:** (UI converge)

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

Here is our  $L_p$  convergence Thm which effectively shows

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

**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  $E|X_n|^p \rightarrow E|X|^p < \infty$

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

Proof:

(9)

(i)  $\Rightarrow$  (ii)

We already know  $X_n \xrightarrow{P} X$  by Markov's neg.

$X \in L_p$  since  $L_p$  is closed. Finally

by  $\| \|X\|_p - \|X_n\|_p \| \leq d_p(X, X_n) \rightarrow 0$

we have

$$E|X_n|^p \rightarrow E|X|^p < \infty$$

(ii)  $\Rightarrow$  (i).

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

$$0 \leq |X_n - X|^p \leq 2^p \left( \frac{1}{2}|X_n|^p + \frac{1}{2}|X|^p \right) =: Y$$

$\downarrow P$                      $\downarrow P$   
 $0$                              $2^p|X|^p$           $=: Y$   
 by continuous  
 mapping since  
 $X_n \xrightarrow{P} X \rightarrow 0$

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

(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  $E|X_n|^p \rightarrow E|X|^p < \infty$  so that  $X_n, X \in L_p$  for suff large  $n$

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

(iii)  $\Rightarrow$  (i):

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

QED

(10)

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

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

$\therefore$  we can define an inner product on  $L_2$  defined as

$$\langle X, Y \rangle := E(XY)$$

Much of what statistics 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

$$\langle X, Y \rangle = \text{cov}(X, Y)$$

$$\langle X, X \rangle = \|X\|_2^2 = \text{var}(X)$$

$$\|X\|_2 = \text{sd}(X).$$

Basic Properties of  $\langle \cdot, \cdot \rangle$ :

$\forall X, Y \in L_2(\mathcal{S}, \mathcal{F}, P)$

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

(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)  $|\langle X, Y \rangle| \leq \|X\|_2 \|Y\|_2$

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

which is true since

$$|\langle X_n, Y \rangle - \langle X, Y \rangle| = |\langle X_n - X, Y \rangle|$$

$$\leq \|X_n - X\|_2 \|Y\|_2$$

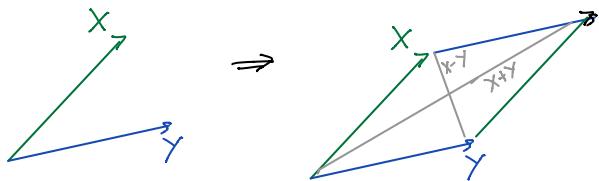
(7)  $\|X + Y\|_2^2 = \|X\|_2^2 + 2\langle X, Y \rangle + \|Y\|_2^2$

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

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.

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

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

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

(11) BLP (12)

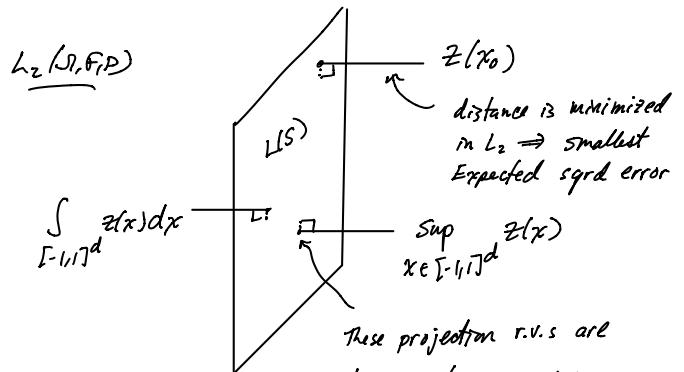
Now "Best linear prediction" of some unobserved  $L_2(\Omega, \mathcal{F}, P)$  r.v. simply given by projection:

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

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



These projection r.v.s are known to us since they are limits of linear combinations of our observations  $Z(x) : x \in [0, 1]^d$

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  $\theta_S Y \in S$  s.t.

$$\|Y - P_S Y\|_2 = \inf_{X \in S} \|Y - X\|_2.$$

Moreover PSY is characterized by the following two properties

- (1)  $\theta_s y \in S$

(2)  $\underbrace{(y - \theta_s y)}_{\text{prediction residual.}} \perp x \quad \forall x \in S$

Proof:

(Find  $\varnothing_S Y$ ) Let  $X_n \in S$  s.t.

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

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

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

$$\therefore \|\chi_n - \chi_m\|_2^2 = I_{nm} - \underbrace{\|(x_n - y) + (x_m - y)\|_2}_{2\left(\frac{x_n+x_m}{2} - y\right)}$$

$$= I_{nm} - 4 \left\| \underbrace{\frac{x_n + x_m}{2} - y}_c \right\|_2^2$$

ES by linearity

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

$\rightarrow 0$  as  $n, m \rightarrow \infty$

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

$$\exists P_S Y \in L_2 \quad \text{s.t.} \quad \underbrace{X_n}_{\in S} \xrightarrow{\mathcal{L}_2} \underbrace{P_S Y}_{\in S} \quad \text{since } S \text{ is closed}$$

Also, for this PSY we have

$$\inf_{X \in S} \|X - Y\|_2 \leq \|\Theta_S Y - Y\|_2$$

$$\leq \underbrace{\|P_S Y - X_n\|_2}_{\rightarrow 0} + \underbrace{\|X_n - Y\|_2}_{\rightarrow \inf_{X \in S} \|X - Y\|_2}$$

$$\therefore \inf_{X \in S} \|X - Y\|_2 = \|\Theta_S Y - Y\|_2$$

(Show  $\theta_{SY}$  is unique P-a.e)

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

Again by the Parallelogram Thm

$$\|(\chi_0 - \gamma) + (\beta_s \gamma - \gamma)\|_2^2 + \|\chi_0 - \beta_s \gamma\|_2^2$$

$$= \underbrace{2\|\chi_0 - \gamma\|_2^2}_{2\inf^2} + \underbrace{2\|\rho_s \gamma - \gamma\|_2^2}_{2\inf^2}$$

$$\begin{aligned} \therefore \|X - P_S Y\|_2^2 &\leq 4 \inf_{X \in S} \|X - Y\|_2^2 = \|2\left(\underbrace{\frac{X_0 + P_S Y}{2} - Y}_{\in S}\right)\|_2^2 \\ &\leq 4 \inf_{X \in S} \|X - Y\|_2^2 = 0 \end{aligned}$$

$$\therefore X_0 = P_S Y.$$

(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:

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

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

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

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

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

$\forall X \in S$  we have

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

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

$$= \|W - Y\|_2^2$$

$\therefore W = P_S Y$  Since  $P_S Y$  is the unique such v.v.

GED

The next theorem shows that to compute a projection you define coordinates aligned with the space your projecting to

Theorem: (projection in coordinates)

Suppose  $X_1, X_2, \dots \in L_2(\Omega, \mathcal{F}, P)$  are orthonormal and let

$$S = \overline{\text{span}}\{X_n : n \in \mathbb{N}\} = \begin{pmatrix} \text{The collection of } L_2 \\ \text{limits of finite} \\ \text{linear combinations} \\ \text{of the } X_n's \end{pmatrix}$$

Then  $S$  is a closed linear subset of  $L_2$  &  $\forall Y \in L_2(\Omega, \mathcal{F}, P)$  the following holds

$$\begin{aligned} \underbrace{P_S Y}_{\substack{\text{Projection} \\ \text{of } Y \text{ onto } S}} &\stackrel{L_2}{=} \sum_{n=1}^{\infty} \langle X_n, Y \rangle X_n \\ &\text{This means} \\ &\sum_{n=1}^N \langle X_n, Y \rangle X_n \xrightarrow[N \rightarrow \infty]{L_2} P_S Y \end{aligned}$$

*Proof:*

( $S$  is closed and linear):

$$W, z \in S \Rightarrow \begin{aligned} W_n &\xrightarrow{L_2} W & \text{where } w_n, z_n \text{ is a finite} \\ z_n &\xrightarrow{L_2} z & \text{linear comb of the } x_i's \\ \Rightarrow \|aw_n + bz_n - (aw + bz)\|_2 &= |a|\|w_n - w\|_2 + |b|\|z_n - z\|_2 \\ &\quad \underbrace{\qquad\qquad\qquad}_{\rightarrow 0} \\ \Rightarrow aw + bz &\in S \end{aligned}$$

$\therefore S$  is linear

To see that  $S$  is closed let

$$z_n \xrightarrow{L_2} z$$

where  $z_n \in S$ . For each  $n$  let  $\hat{z}_n$  be a finite linear combination of the  $x_i$ 's s.t.

$$\|\hat{z}_n - z_n\|_2 = \frac{1}{n}$$

$$\therefore \|z - \hat{z}_n\|_2 \leq \|z - z_n\|_2 + \frac{1}{n} \rightarrow 0 \text{ as } n \rightarrow \infty$$

$\therefore z \in S$  by definition.

( $\sum_{i=1}^{\infty} \langle x_i, y \rangle x_i$  exists as a infinite sum)

Let  $S_n := \overline{\text{span}}\{x_1, \dots, x_n\}$

Notice that  $P_{S_n} y = \sum_{i=1}^n \langle x_i, y \rangle x_i$  since

$$\langle y - P_{S_n} y, x_j \rangle = \langle y, x_j \rangle - \langle x_j, y \rangle = 0$$

$\forall j = 1, \dots, n$ . Also notice that  $P_{S_n} y$  decreases length in that

$$\|y\|_2^2 = \|y - P_{S_n} y\|_2^2 + \|P_{S_n} y\|_2^2$$

$$\geq \|P_{S_n} y\|_2^2$$

$$\therefore \sum_{i=1}^n \langle x_i, y \rangle^2 = \|P_{S_n} y\|_2^2 \leq \|y\|_2^2 < \infty$$

(17)

$$\therefore \sum_{i=1}^{\infty} \langle x_i, y \rangle^2 < \infty$$

$\therefore \left\{ \sum_{i=1}^n \langle x_i, y \rangle x_i \right\}_{n \geq 1}$  is Cauchy in  $L_2$

and has a  $L_2$  limit, denoted  $\sum_{i=1}^{\infty} \langle x_i, y \rangle x_i$ .

(Show  $P_S y = \sum_{i=1}^{\infty} \langle x_i, y \rangle x_i$ ):

$$\text{Since } \sum_{i=1}^n \langle x_i, y \rangle x_i \xrightarrow[n \rightarrow \infty]{L_2} \sum_{i=1}^{\infty} \langle x_i, y \rangle x_i \\ = P_{S_n} y$$

we have that  $\forall j \in \mathbb{N}$

$$\langle P_{S_n} y, x_j \rangle \xrightarrow{n \rightarrow \infty} \left\langle \sum_{i=1}^{\infty} \langle x_i, y \rangle x_i, x_j \right\rangle$$

$$\therefore \langle y - P_{S_n} y, x_j \rangle \xrightarrow{n \rightarrow \infty} \underbrace{\langle y - \sum_{i=1}^{\infty} \langle x_i, y \rangle x_i, x_j \rangle}_{\text{This is eventually zero for large enough } n} \Rightarrow \text{This is } 0$$

$$\therefore P_S y = \sum_{i=1}^{\infty} \langle x_i, y \rangle x_i \text{ since } y - P_S y \perp x \quad \forall x \in S.$$

QED.

Definition:  $\{x_i : i \in \mathbb{N}\} \subset L_2(\Omega, \mathcal{F}, P)$  is an orthonormal basis (ONB) if finite linear combinations of the  $x_i$ 's are dense in  $L_2(\Omega, \mathcal{F}, P)$ .

### Theorem: (characterizing a ONB)

(19)

If  $\{X_n\}_{n \geq 1} \subset L_2(\Omega, \mathcal{F}, P)$  are orthonormal then the following are equivalent:

(i)  $\{X_n\}_{n \geq 1}$  is a ONB

(ii)  $Y = \sum_{i=1}^{\infty} \langle X_i, Y \rangle X_i \quad \forall Y \in L_2(\Omega, \mathcal{F}, P)$

(iii)  $\langle Y, z \rangle = \sum_{i=1}^{\infty} a_i b_i \quad \forall Y, z \in L_2(\Omega, \mathcal{F}, P)$

where  $a_i := \langle X_i, Y \rangle$  &  $b_i := \langle X_i, z \rangle$ .

(iv)  $\|Y\|_2^2 = \sum_{i=1}^{\infty} a_i^2 \quad \forall Y \in L_2(\Omega, \mathcal{F}, P)$

where  $a_i := \langle X_i, Y \rangle$ .

Proof:

(ii)  $\Rightarrow$  (i): Trivial

(i)  $\Rightarrow$  (ii): Almost trivial.

If  $S := \overline{\text{span}} \{X_i : i \in \mathbb{N}\}$  then

$$P_S Y = \sum_{i=1}^{\infty} \langle X_i, Y \rangle X_i \quad \text{by "Comparing a projection Thm"}$$

$\Downarrow$

Since (i)

implies  $S = L_2(\Omega, \mathcal{F}, P)$ .

(ii)  $\Rightarrow$  (iii): Set  $a_i := \langle X_i, Y \rangle$ ,  $b_i := \langle X_i, z \rangle$  and notice

$$\begin{aligned} \langle Y, z \rangle &= \left\langle \sum_{i=1}^{\infty} a_i X_i, \sum_{i=1}^{\infty} b_i X_i \right\rangle \\ &= \lim_n \left\langle \sum_{i=1}^n a_i X_i, \sum_{i=1}^{\infty} b_i X_i \right\rangle \\ &= \lim_n \lim_m \underbrace{\left\langle \sum_{i=1}^n a_i X_i, \sum_{i=1}^m b_i X_i \right\rangle}_{\sum_{i=1}^n a_i b_i} \end{aligned}$$

(iii)  $\Rightarrow$  (iv): Trivial

(iv)  $\Rightarrow$  (ii):

(20)

Let  $S := \overline{\text{span}} \{X_i : i \in \mathbb{N}\}$ .

Since we know  $P_S Y = \sum_{i=1}^{\infty} \langle X_i, Y \rangle X_i$  by

"Projection in coordinates Thm" it will be sufficient to show

$$\|P_S Y - Y\|_2^2 = 0$$

$$\text{since } \|Y\|_2^2 = \|Y - P_S Y\|_2^2 + \|P_S Y\|_2^2$$

$$\underbrace{\sum_{i=1}^{\infty} \langle X_i, Y \rangle^2}_{\text{by (iv)}} \quad \underbrace{\sum_{i=1}^{\infty} \langle X_i, Y \rangle^2}$$

we therefore have  $\|Y - P_S Y\|_2^2$  as was to be shown

QED

Our last Thm on projections shows that the ordering of the ONB is irrelevant.

### Theorem: (permuting coordinates)

If  $\{X_n\}_{n \geq 1} \subset L_2(\Omega, \mathcal{F}, P)$  is an orthonormal collection and  $Y \in L_2(\Omega, \mathcal{F}, P)$  then

$$\sum_{i=1}^{\infty} \langle X_i, Y \rangle X_i = \sum_{i=1}^{\infty} \langle X_{\pi(i)}, Y \rangle X_{\pi(i)}$$

for any permutation  $\pi: \mathbb{N} \rightarrow \mathbb{N}$ .

Proof:

Let  $Z = \sum_{i=1}^{\infty} \langle X_i, Y \rangle X_i$  and

$$Z_{\pi} = \sum_{i=1}^{\infty} \langle X_{\pi(i)}, Y \rangle X_{\pi(i)}.$$

Now

$$\begin{aligned} \langle Z, Z_{\pi} \rangle &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \langle X_i, Y \rangle \langle X_{\pi(j)}, Y \rangle \underbrace{\langle X_i, X_{\pi(j)} \rangle}_{\begin{cases} 1 & \text{if } i = \pi(j) \\ 0 & \text{o.w.} \end{cases}} \\ &= \sum_{i=1}^{\infty} \langle X_i, Y \rangle^2 \\ &= \|Z\|_2^2 \end{aligned}$$

$$\begin{aligned} \therefore \|z - z_{\pi}\|_2^2 &= \|z\|_2^2 + \|z_{\pi}\|_2^2 - 2\langle z, z_{\pi} \rangle \\ &= \sum_{i=1}^{\infty} \langle x_{\pi(i)}, y \rangle^2 - \sum_{i=1}^{\infty} \langle x_i, y \rangle^2 \\ &= 0 \quad \text{since positive} \\ &\quad \text{sums can be arbitrarily} \\ &\quad \text{re-ordered.} \end{aligned} \tag{21}$$

QED

Remark: Notice that whenever  $\mathcal{F}$  is countably generated,  $L_2(\Omega, \mathcal{F}, P)$  is separable so there exists a countable dense subset  $\{X_n\}_{n \geq 1}$ .

In this case one can use Gram-Schmidt to construct an ONB  $\{X_n\}_{n \geq 1}$  as follows:

$$X_1 := \frac{Y_1}{\|Y_1\|_2}$$

$$X_2 := \frac{Y_2 - \langle X_1, Y_2 \rangle X_1}{\|Y_2 - \langle X_1, Y_2 \rangle X_1\|_2} \leftarrow \text{project out}$$

$$X_n := \frac{Y_n - \sum_{i=1}^{n-1} \langle X_i, Y_n \rangle X_i}{\|Y_n - \sum_{i=1}^{n-1} \langle X_i, Y_n \rangle X_i\|_2} \leftarrow \begin{array}{l} \text{project out} \\ \{X_i\}_{i=1}^{n-1} \\ \text{normalize} \end{array}$$

Now each  $X_n$  is a finite linear comb of the  $X_i$ 's so  $\{X_n\}_{n \geq 1}$  are dense in  $L_2$ . If we identify r.v.'s with the equivalence classes of P-a.e. modifications then  $L_2(\Omega, \mathcal{F}, P)$  becomes a **Hilbert Space** (complete, separable, linear vector space with pos. def. inner product).

### Projection Example

(22)

$$\Omega = [-1, 1]$$

$$\mathcal{F} = \mathcal{B}([-1, 1])$$

$P = \frac{1}{2} \times \text{Lebesgue measure on } [-1, 1]$

$$S = \{X \in L_2(\Omega, \mathcal{F}, P) : X(w) = X(-w) \text{ P-a.e.}\}$$

Clearly  $S$  is linear. It is also closed

$$\begin{aligned} \text{Since } \underbrace{X_n}_{\in S} &\xrightarrow{L_2} X \Rightarrow X_n \xrightarrow{P} X \\ &\Rightarrow X_{n_k} \xrightarrow{\text{a.e.}} X \text{ for some sub-seq } n_k \\ &\Rightarrow X(w) = \lim_{k \rightarrow \infty} X_{n_k}(w) \\ &\stackrel{\text{a.e.}}{=} \lim_{k \rightarrow \infty} X_{n_k}(-w) \\ &\stackrel{\text{a.e.}}{=} X(-w) \end{aligned}$$

For any  $Y \in L_2(\Omega, \mathcal{F}, P)$  let's find  $P_S Y$ .

Technique: Guess the answer and show it is orthogonal to  $S$ .

Here is the guess:

$$P_S Y = \frac{Y(w) + Y(-w)}{2}$$

To verify let  $X \in S$  and notice

$$\langle Y - P_S Y, X \rangle = E[(Y - P_S Y) X]$$

$$= \int_{[-1, 1]} (Y(w) - \frac{Y(w) + Y(-w)}{2}) X(w) dP(w)$$

$$= \frac{1}{2} \int_{[-1, 1]} (Y(w) - Y(-w)) X(w) dP(w)$$

odd symmetry  
about 0

$$= 0$$

$$\therefore \text{Indeed } P_S Y = \frac{Y(w) + Y(-w)}{2}.$$

Projection application to  
Gaussian random field prediction

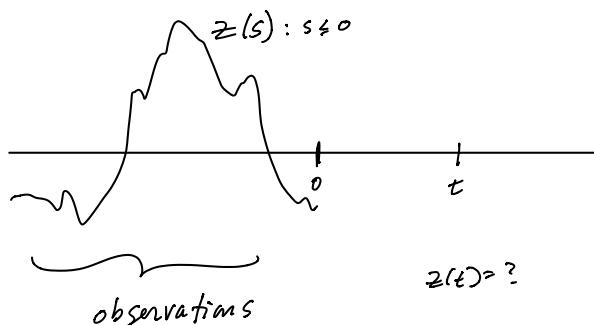
(23)

Let  $\{z(t) : t \in \mathbb{R}\}$  be a Gaussian random field (GRF) s.t.  $\forall t, s \in \mathbb{R}$

$$\begin{aligned} E(z(t)) &= 0 \\ \text{cov}(z(t), z(s)) &= e^{-|t-s|} \end{aligned} \quad \left. \begin{array}{l} \text{specifies} \\ \text{the f.d.d.s.} \end{array} \right\}$$

Note part of the requirement of a GRF is that all the random variables  $z(t)$ , for  $t \in \mathbb{R}$ , must be defined on the same probability space  $(\Omega, \mathcal{F}, P)$ .

Suppose we observe  $z(s)$ ,  $\forall s \leq 0$  & want to predict  $z(t)$  for some  $t > 0$ .



Now, we will see later that

$$E(z(t) | z(s), s \leq 0) = P_s z(t)$$

To find  $P_s z(t)$  let's guess that  $P_s z(t) = a_t z(0)$  and prove the residuals are orthogonal to  $\overline{\text{span}} \{z(s) : s \geq 0\}$ .  
Set of all linear  $L_2$  quantities from  $z(s), s \leq 0$ .

Note that by linearity & continuity of  $\langle \cdot, \cdot \rangle$  w.r.t  $L_2$  limits it is sufficient to show  $\langle z(t) - a_t z(0), z(s) \rangle = 0 \quad \forall s \leq 0$ .

For  $t > 0$  &  $s \leq 0$  we have

$$\langle z(t) - a_t z(0), z(s) \rangle$$

$$= e^{-|t-s|} - a_t e^{-|s|}$$

$$= e^{-(t-s)} - a_t e^s \quad \text{since } s \leq 0 < t$$

$$= 0 \quad \text{iff } a_t := e^{-t}$$

$$\therefore E(z(t) | z(s), s \leq 0) = e^{-t} z(0)$$

and

$$\text{var}(z(t) | z(s), s \leq 0)$$

$$= E((z(t) - e^{-t} z(0))^2 | z(s), s \leq 0)$$

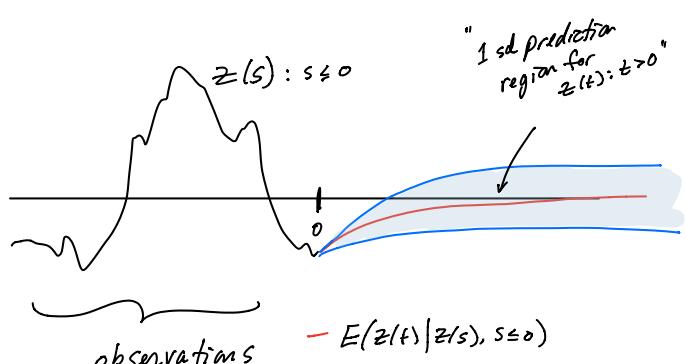
Since these are orthogonal in  $L_2$ , they are uncorrelated.  
Since they are Gaussian they are then independent

$$= \text{var}(z(t) - e^{-t} z(0))$$

marginal residual

$$= e^0 - 2e^{-t} e^{-|t|} + e^{2t} e^0$$

$$= 1 - e^{-2t}$$



$$- E(z(t) | z(s), s \leq 0)$$

$$- E(z(t) | z(s), s \leq 0)$$

$$\pm \text{sd}(z(t) | z(s), s \leq 0)$$

## Riesz representation

(25)

The Riesz representation theorem will be important for proving the existence of  $\frac{dP}{dQ}$  for two probability measures  $P, Q$ .

To moderate recall that if

$X_n, X \in L_2(\Omega, \mathcal{F}, P)$  then for any  $Y \in L_2(\Omega, \mathcal{F}, P)$

$$X_n \xrightarrow{L_2} X \Rightarrow \langle Y, X_n \rangle \rightarrow \langle Y, X \rangle$$

$\therefore$  the map  $\langle Y, \cdot \rangle: L_2(\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}$  is a continuous linear functional on  $L_2$ .

Riesz says that all such functionals are in the form  $\langle Y, \cdot \rangle$  for some  $Y \in L_2$ .

### Theorem (Riesz for $L_2$ ):

Let  $f: L_2(\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}$  be linear and continuous (w.r.t  $L_2$  metric).

Then  $\exists$  a  $P$ -a.e. unique  $Y \in L_2(\Omega, \mathcal{F}, P)$  s.t.

$$f(X) = \langle Y, X \rangle \quad \forall X \in L_2(\Omega, \mathcal{F}, P).$$

Proof:

The idea is that if such a  $Y$  did exist, then one could Gram-Schmidt to get a ONB:

$$\underbrace{cY, X_1, X_2, \dots}_{\text{This spans } N^\perp}$$

This spans  $N^\perp$  where  $\dim N^\perp = 1$

Now this spans the null space of  $f$ ,  $N := \{X : f(X) = 0\}$

So let's find  $Y \in N^\perp$  with the desired property.

Step 1: choose any  $Y \notin N$  s.t.  $\underbrace{Y \neq 0}_{\text{if no such } Y \text{ then } f(x)=0 \ \forall X \text{ & setting } Y=0 \text{ gives the result.}}$  P-a.e.

Step 2: Project out  $N$

$$\text{i.e. } \underbrace{Y := Y - P_N Y}_{\text{new}}$$

Note  $N$  is closed & linear by assumptions on  $f$  so  $P_N Y$  exists.

Step 3: scale  $Y$  by a constant so it satisfies  $f(Y) = \|Y\|_2^2$ .

Now I can project any  $X \in L_2$  to  $\text{span } Y := \{cY : c \in \mathbb{R}\}$  by

$$P_{\text{span } Y} X = \left\langle \frac{Y}{\|Y\|_2}, X \right\rangle \frac{Y}{\|Y\|_2}$$

and is characterized by

$$\left\langle X - P_{\text{span } Y} X, Y \right\rangle = 0 \quad \forall X$$

$$\left\langle X - \frac{\langle Y, X \rangle}{\|Y\|_2} Y, Y \right\rangle$$

If it was the case that  $f(X) = \langle Y, X \rangle$  then it should satisfy

$$f\left(X - \frac{f(X)}{\|Y\|_2^2} Y\right) = 0$$

Indeed

$$f\left(X - \frac{f(X)}{\|Y\|_2^2} Y\right) = f(X) - \frac{f(X)}{\|Y\|_2^2} f(Y) = 0$$

$$\therefore X - \frac{f(X)}{\|Y\|_2^2} Y \in N \leftarrow \text{the null space of } f \quad (27)$$

$$\therefore \left\langle X - \frac{f(X)}{\|Y\|_2^2} Y, Y \right\rangle = 0 \quad \text{since we projected } N \text{ out of } Y.$$

$$\therefore \left\langle X, Y \right\rangle - \frac{f(X)}{\|Y\|_2^2} \left\langle Y, Y \right\rangle = 0$$

$$\therefore f(X) = \left\langle Y, X \right\rangle \quad \forall X \in L_2(\Omega, \mathcal{F}, P).$$

To show uniqueness let  $\tilde{Y} \in L_2$  s.t.

$$\left\langle \tilde{Y}, X \right\rangle = f(X) = \left\langle Y, X \right\rangle \quad \forall X \in L_2$$

$$\therefore \left\langle Y - \tilde{Y}, X \right\rangle = 0 \quad \forall X \in L_2$$

$$\therefore \left\langle Y - \tilde{Y}, \underbrace{Y - \tilde{Y}}_{\in L_2} \right\rangle = 0$$

$$\therefore Y = \tilde{Y} \quad P-a.e.$$

QED