Hilbert spaces. Conditional Expectation. Prediction

Bojan Basrak
University of Zagreb

II HILBERT SPACES. CONDITIONAL EXPECTATION & PREDICTION

Given observations $X_1, \dots, X_n$, we can try to predict X_{n+1}, e.g. by minimizing

$$\min_{g} E|X_{n+1} - g(X_1, \dots, X_n)|^2$$

or

$$\min_{\alpha_i} E|X_{n+1} - \sum_{1}^{n} \alpha_i X_i|^2$$

Recall: H is a Hilbert space if it is a complete inner-product space.

THEOREM 1 (on projection)

If M is a closed subspace of Hilbert space H, & $x \in H$ then

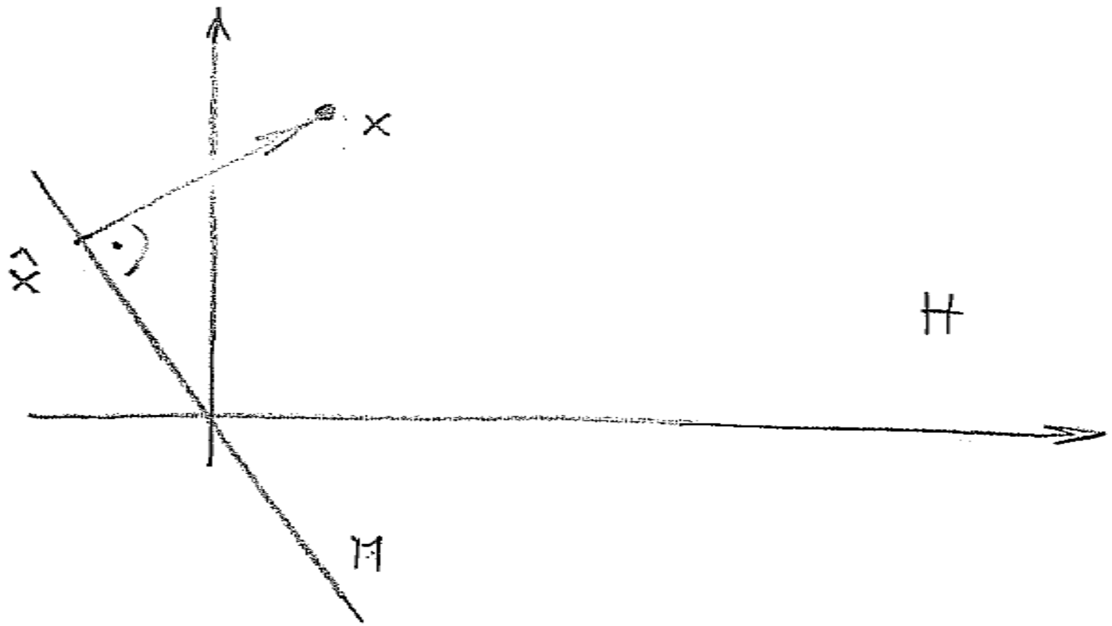
i) there is an unique $\hat{x} \in M$ s.t.

$$\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|$$

ii) $\hat{x} \in M$ & $\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|$

$\iff$ $\hat{x} \in M$ & $(x - \hat{x}) \in M^{\perp}$

$M^\perp$ = orthogonal complement of M



Mapping $x \longmapsto \hat{x} =: \Pi_M(x)$ is called orthogonal projection.

We know:

- Π_M is a linear operator
- $\| \Pi_M x \| \leq \| x \| \qquad \forall\, x \in H$
- $\Pi_M^2 = \Pi_M$
- $M_1 < M_2 \qquad$ (M_1 is closed subsp of M_2)

  $\Rightarrow \quad \Pi_{M_1} \Pi_{M_2} x = \Pi_{M_1} x$
- $M_1 \perp M_2$, closed subspaces of H

  $\Rightarrow \quad \Pi_{M_1+M_2} x = \Pi_{M_1} x + \Pi_{M_2} x$

HILBERT SPACE $L_2(\Omega, \mathcal{F}, \mathbb{P})$

Recall : elements of this Hilbert space are equivalence classes of r.v.'s

although we write

$$X \in L_2(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{if} \quad EX^2 < \infty$$

$$X \sim Y \quad \text{if} \quad P(X \neq Y) = 0$$

Sometimes even complex random variables are considered

$Z = X + iY$, where (X, Y) is r.vector on $(\Omega, \mathcal{F}, \mathbb{P})$

define :

$$EZ = EX + iEY$$

$$\operatorname{Var} Z = E|Z - EZ|^2$$

$$\operatorname{Cov}(Z_1, Z_2) = E(Z_1 - EZ_1)\overline{(Z_2 - EZ_2)}$$

EXE 1) Show

i) expectation of complex r.v's is linear

ii) $\operatorname{Var}(\alpha Z) = |\alpha|^2 \operatorname{Var} Z$, $\alpha \in \mathbb{C}$

iii) $\operatorname{Cov}(Z_1, Z_2) = EZ_1\overline{Z_2} - EZ_1 E\overline{Z_2}$

Corresponding norm

$$\|X\| = \sqrt{E|X|^2} = \sqrt{\langle X, X \rangle}$$

Convergence in L_2:

$$X_n \xrightarrow{L_2} X \quad \text{means} \quad E|X_n - X|^2 \longrightarrow 0$$

Cauchy Schwarz inequality

$$|\langle x, y \rangle|^2 \le \|X\|^2 \|y\|^2$$

i.e. $$|EX\bar{y}| \le \sqrt{E|X|^2 \cdot E|y|^2} \qquad (1)$$

EXAMPLE 1

If $X_n \xrightarrow{L_2} X$, $Y_n \xrightarrow{L_2} Y \Rightarrow$

$$\langle X_n, Y_n \rangle \longrightarrow \langle X, Y \rangle \text{ in } \mathbb{R} \text{ or } \mathbb{C}$$

I.e. inner-product is continuous

EXE 2> Use (1) to show

$$X_n \xrightarrow{L_2} X \Rightarrow X_n \xrightarrow{L_1} X$$

EXE 3> Show $X, Y \in L_2$

$$s.d.(X+Y) \le s.d.(X) + s.d.(Y)$$

CONDITIONAL EXPECTATION

Suppose $\mathcal{F}_0$ is sub σ-algebra of $\mathcal{F}$

$\Downarrow$

$\mathcal{F}_0$ measurable r.v.'s $Y \in L_2(\Omega, \mathcal{F}, \mathbb{P})$ form a closed subspace of $L_2(\Omega, \mathcal{F}, \mathbb{P})$

i.e. $L_2(\Omega, \mathcal{F}_0, \mathbb{P}) \leq L_2(\Omega, \mathcal{F}, \mathbb{P})$

$\boxed{\text{DEF}}$ Projection of $X \in L_2(\Omega, \mathcal{F}, \mathbb{P})$ on $L_2(\Omega, \mathcal{F}_0, \mathbb{P})$ is called <u>conditional expectation</u> of X w.r.t. $\mathcal{F}_0$, notation $E(X \mid \mathcal{F}_0)$.

$\boxed{\text{DEF}}^2$ Conditional expectation of nonnegative or integrable r.v. X w.r.t. $\mathcal{F}_0$ is a $\mathcal{F}_0$-measurable r.v. (notation $E(X|\mathcal{F}_0)$) s.t.

$$\int_A E(X|\mathcal{F}_0)\,dP = \int_A X\,dP \qquad \forall A \in \mathcal{F}_0$$

i.e.
$$E\left[(X|\mathcal{F}_0)\cdot 1_A\right] = E\left[X 1_A\right] \qquad \forall A \in \mathcal{F}_0$$

REM 2nd def is more common; by it $E(X|\mathcal{F}_0)$ is a r.v., by 1st def it is an element of L_2 (!?!?).

- By def. 2, cond. expectation is not unique, only $\mathbb{P}|\mathcal{F}_0$ - a.s. unique.
- Existence of $E(X|\mathcal{F}_0)$ in def 2. is guaranteed by Radon - Nikodymov theorem.
- For $X \in L_2$, two definitions are essentially equivalent

$\boxed{\text{DEF}}$ For nonnegative or integrable X.

$$E(X|Y) := E(X|\sigma(Y)),$$

where Y is an arbitrary r. vector on the same pr. space.

EXAMPLE 2

i) $\mathcal{F}_0 = \{\emptyset, \Omega\}$, clearly $E(X|\mathcal{F}_0) = EX \in \mathbb{R}$ is cond. expect. by def. 2.

ii) If X is $\mathcal{F}_0$-measurable then by def. 2

$$E(X|\mathcal{F}_0) = EX$$

REM Increasing seq. of σ-algebra is used to model increasing amount of information collected over time. Such a sequence of σ-algebras $(\mathcal{F}_t)$ is called filtration.

THEOREM 2 (properties of cond. exp.)

For X, Y integr. r.v's on $(\Omega, \mathcal{F}, \mathbb{P})$ & $\mathcal{G} \subseteq \mathcal{F}$ σ-algebra, it holds

i) $E(\alpha X + \beta Y \mid \mathcal{G}) = \alpha E(X \mid \mathcal{G}) + \beta E(Y \mid \mathcal{G})$ a.s.

ii) Z $\mathcal{G}$-measurable $\Rightarrow$

$$E(ZX \mid \mathcal{G}) = Z E(X \mid \mathcal{G}) \quad \text{a.s.}$$

(if $Z, X \in L_2$ e.g.)

iii) $X \geq 0$ a.s $\Rightarrow E(X|\mathcal{G}) \geq 0$ a.s.

iv) if $\mathcal{G}_0 \subseteq \mathcal{G} \subseteq \mathcal{F}$ σ-algebra

$$E\left[E(X|\mathcal{G})|\mathcal{G}_0\right] = E\left[X|\mathcal{G}_0\right] \quad \text{a.s.}$$

in particular

$$E(E(X|\mathcal{G})) = EX$$

(when $\mathcal{G}_0 = \{\emptyset, \Omega\}$)

LEMMA 3 (Dudley)

For r.v.'s $(Y_\alpha : \alpha \in A)$, if $X \in \sigma(Y_\alpha : \alpha \in A)$

then

i) $|A| = k < \infty \Rightarrow \exists$ measurable $g : \mathbb{R}^k \to \mathbb{R}$

$$X = g(Y_1, \ldots, Y_k)$$

ii) $|A| = +\infty \Rightarrow \exists$ countable set of indices $\{\alpha_n\}_n \in A$

& measurable $g : \mathbb{R}^\infty \longrightarrow \mathbb{R}$ s.t.

$$X = g(Y_{\alpha_1}, Y_{\alpha_2}, \ldots)$$

Thus, since $E(X|Y)$ is $\sigma(Y)$- measurable

$\exists\ g$ s.t. $E(X|Y) = g(Y)$

& we write

$$g(y) =: E(X|Y=y)$$

although this does not have usual interpretation if $P(Y=y)=0$.

LINEAR PREDICTORS

Assume: (X_t) is weakly stationary sequence with mean zero.

DEF Suppose $EX_t = 0 \ \forall t$. The best linear predictor for X_{n+1} in terms of $X_1, \dots, X_n$ is the linear combination

$$Z = c_1 X_1 + \dots + c_n X_n \quad \text{s.t.}$$

$$\inf_{y \in \operatorname{span}(X_1, \dots, X_n)} E|X_{n+1} - y|^2 = E|X_{n+1} - Z|^2 \quad \leftarrow \boxed{\text{SQUARE PREDICTION ERROR}} \quad (2)$$

Observe : $\mathrm{span}(X_1, \ldots, X_n) =: M_n$ is closed subspace of $L_2(\Omega, \mathcal{F}, \mathbb{P})$. By projection theorem $\Rightarrow$

$$\exists!\ Z \in M_n \text{ s.t. } (2) \text{ holds}$$

We denote projection of L_2 on M_n by Π_n, i.e.

$$\Pi_n X_{n+1} = \varphi_1 X_n + \cdots + \varphi_n X_1$$

REM. We could the same & calculate $\Pi_n W$ for $W = X_{n+h}$ for instance.

- φ_i's depend on n, they should be called $\varphi_{n1}, \varphi_{n2}, \ldots, \varphi_{nn}$.

By projection thm again Z in (2) is characterized

by $Z \in M$, $X_{n+1} - Z \perp M$

$$\Leftrightarrow \quad \langle X_{n+1} - Z, X_i \rangle = 0 \quad , i = 1, .., n \qquad (*)$$

$$\Leftrightarrow \quad \langle X_{n+1}, X_i \rangle = \langle \ell_1 X_n + \cdots + \ell_n X_1, X_i \rangle$$

$i = 1, \ldots, n$

$$= \ell_1 \langle X_n, X_i \rangle + \cdots + \ell_n \langle X_1, X_i \rangle$$

$i = 1, .., n \qquad (**)$

But $\langle X_j, X_i \rangle = E X_j X_i = Cov(X_j, X_i)$

$$= \gamma_X(j-i) \text{ , so}$$

$(**) \Longleftrightarrow$

$$\gamma(n+1-i) = \ell_1 \gamma(n-i) + \cdots + \ell_n \gamma(1-i) \qquad i = 1, \ldots, n$$

or

$$\Gamma_n \begin{pmatrix} \ell_1 \\ \\ \ell_n \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) & & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ddots & \\ & \ddots & & \\ \gamma(n-1) & \cdots & & \gamma(0) \end{pmatrix} \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_n \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix} \qquad (***)$$

Any of the systems of equations above is called the prediction equations in t.s.a.

We know that one solution to (***) exists, are there any other solutions ?

Uniqueness $\Leftrightarrow$ Γ_n is regular matrix

$\Leftrightarrow$ Γ_n is positive definite matrix

↳ there can be more than one solution, i.e. linear combination which minimizes square prediction error

↳ the best lin. predictor can be found from autocov. function γ, however γ has to be estimated, together with predictors

Note:

$$E|X_{n+1} - \Pi_n X_{n+1}|^2 = \gamma_X(0) - \left\langle \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_n \end{pmatrix}^\tau, \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix} \right\rangle$$

EXE 4) Show: the best l.m. predictor for X_{n+h} solves

$$\Gamma_n \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \gamma(h) \\ \vdots \\ \gamma(n+h-1) \end{pmatrix}$$

$\alpha_1 X_n + \cdots + \alpha_n X_1$ is $\Pi_n X_{n+h}$; show

$$E|X_{n+h} - \Pi_n X_{n+h}|^2 = \gamma(0) - (\alpha_1, \ldots, \alpha_n)' (\gamma(h), \ldots, \gamma(n+h-1)).$$

EXAMPLE 2 (AR(1) process)

For $Z_t \sim WN(0,\sigma^2)$, (X_t) satisfies (suppose)

$$X_t = \varphi X_{t-1} + Z_t, \quad t \in \mathbb{Z}, \quad |\varphi| < 1$$

Since $X_t = \sum_{i=0}^{\infty} \varphi^i Z_{t-i}$

X_j & Z_{n+1} are uncorrelated (orthogonal)** for $j \le n$ $\Rightarrow$

$$\Pi_n X_{n+1} = \Pi_n(\varphi X_n) + \underbrace{\Pi_n Z_{n+1}}_{=0}$$

$$= \varphi X_n \qquad (3)$$

EXE 5) Show $\Pi_n X_{n+h} = \varphi^h X_n$ for $h \geq 1$.

Square prediction error in Ex. 2 is

$$E|X_{n+1} - \varphi X_n|^2 = E Z_{n+1}^2 = \sigma^2$$

REM Note (3) is not valid if $|\varphi| > 1$, for ** we have used continuity of inner product.

EXAMPLE 3 (periodic process)

Suppose $X_t = A_1 \cos \nu t + A_2 \sin \nu t$

$\operatorname{Var} A_1 = \operatorname{Var} A_2 = \sigma^2$, $EA_1 = EA_2 = 0$, $\operatorname{Cov}(A_1, A_2) = 0$

CASE OF NON-ZERO EXPECTATION

It makes sense to look for the best lin. predictor in the space

$$M_n' = \mathrm{span}(1, X_1, \dots, X_n)$$

i.e. try to minimize

$$E|X_{n+1} - \varphi_0 - \varphi_1 X_n - \dots - \varphi_n X_1|^2$$

$\boxed{\text{DEF}}$ The best linear predictor for X_{n+1} in terms of $1, X_1, \dots, X_n$ is the projection of X_{n+1} on M_n'.

Note if $EX_t = 0$

$$\Pi_n' := \Pi_{M_n'} = \Pi_n \qquad (p\ [illegible])$$

Therefore if $EX_t = \mu \in R$

$$\Pi_n' X_{n+1} = \mu + \Pi_{\mathrm{span}(X_1-\mu,\dots,X_n-\mu)}(X_{n+1})$$

$$= \mu + \Pi_{\mathrm{span}(X_1-\mu,\dots,X_n-\mu)}(X_{n+1}-\mu)$$

$Y_i := X_i - \mu$ have mean zero $\Rightarrow$ $Y_i \perp \mu$ &

$$\Pi_n' = \Pi_{\mathrm{span}\{1\}} + \Pi_{\mathrm{span}\{Y_1,\dots,Y_n\}}$$

since

$$\mathrm{span}\{1, Y_1, \dots, Y_n\} = \mathrm{span}\{1, X_1, \dots, X_n\}$$

Therefore:

↳ If we have stationary sequence with mean $\mu \neq 0$, we can: 1) subtract the mean

2) then find the best lin. predictor of y_{n+1} in terms of $y_1, \ldots, y_n$, &

3) then add back the estimated mean.

NONLINEAR PREDICTORS

Predictors of the form $f(X_1, .., X_n)$ for X_{n+1} will, of course, have smaller error.

DEF The best predictor for X_{n+1} in terms of $X_1, .., X_n$ is a r.v. $f_n(X_1, \ldots, X_n)$ which minimizes $E|X_{n+1} - f_n(X_1, .., X_n)|^2$ in the class of all measurable functions $f_n : \mathbb{R}^n \to \mathbb{R}$.

By the definition 1 of cond. expectation, the best predictor is

$$E(X_{n+1} \mid X_1, .., X_n)$$

For an arbitrary r.v. W, the best predictor is really

$$E(W \mid X_1, \ldots, X_n)$$

EXAMPLE 4 (ARCH(1) & prediction)

Suppose $|\alpha_1| < 1$, then ARCH(1) has stat. solution &

$$E(X_{n+1} \mid X_t, t \le n) = 0$$

$$E(X_{n+1} \mid X_1, \ldots, X_n) = 0$$

EXAMPLE 5 (AR(1) & prediction)

Assume $|\varphi| < 1$, $(Z_t) \sim \mathrm{IID}(0, \sigma^2)$ then

$$X_t = \sum_0^\infty \varphi^i Z_{t-i}$$

represents a stationary solution to AR(1) equation. Then

$$E(X_{n+1} \mid X_1, \dots, X_n) = \varphi X_n$$

i.e. the best predictor is really the best linear predictor.

PARTIAL AUTOCORRELATION FUNCTION

Assume:

(X_t) is weakly stationary with mean zero

Denote by

$$\Pi_{(2,..,h)} = \text{projection on closed space } \overline{\text{span}}(X_2, \dots, X_h)$$

&

$$\alpha(1) = \text{Corr}(X_2, X_1) = \rho_X(1)$$

$$\alpha(h) = \text{Corr}\left(X_{h+1} - \Pi_{(2,..,h)} X_{h+1},\ X_1 - \Pi_{(2,..,h)} X_1\right)$$

$\boxed{\text{DEF}}$ For such a process (X_t) the sequence / function $\alpha = \alpha_X : \mathbb{N} \to [-1,1]$ introduced above is called the <u>partial autocorrelation function</u> (pacf).

If $EX_t \equiv \mu \neq 0$, pacf is defined as the pacf of $(X_t - \mu)_t$

The value $\alpha(h)$ we can interpret as a correlation between X_t, X_{t+h} when their dependence on intermediate values $X_{t+1}, \ldots, X_{t+h-1}$ has been removed.

EXAMPLE 6 (AR(1) & pacf)

Assume (X_t) is a stationary mean zero & satisfies

$$X_t = \varphi X_{t-1} + Z_t$$

for some $|\varphi| < 1$. Then

$$\begin{aligned}\alpha(1) &= \mathrm{Corr}(X_2, X_1) \\ &= \mathrm{Corr}(\varphi X_1 + Z_2, X_1) \\ &= \varphi\end{aligned}$$

Also $\Pi_{(2,\dots,h)} X_{h+1} = \varphi X_h$ p64.

$$\Pi_{(2,\dots,h)} X_1 = \Pi_{(2,\dots,h)} \left(\frac{1}{\varphi} X_2 - Z_2 \right)$$

$$= \Pi_{(3,\dots,h)} \left(\frac{1}{\varphi} \sum_0^\infty \varphi^i Z_{2-i} - Z_2 \right)$$

REM We say $\alpha(h)$ is pacf at lag h.

LEMA 1

Assume (X_t) is weakly stationary sequence with mean zero, and

$$\Pi_n X_{n+1} = \varphi_1^{(n)} X_n + \cdots + \varphi_n^{(n)} X_1, \quad n \in \mathbb{N}$$

is the best lin. predictor for X_{n+1} in terms of $X_1, \ldots, X_n$ then

$$\alpha_X(n) = \varphi_n^{(n)}$$

EXAMPLE 6 (AR(1) & pacf cont.)

Assume (X_t) is weakly (& strongly) stationary AR(1) process with mean 0, s.t.

$$X_t = \varphi X_{t-1} + Z_t, \quad t \in \mathbb{Z}$$

for $|\varphi| < 1$. The best lin. predictor of X_{h+1} in terms of $X_1, \dots, X_h$ is

$$\Pi_h X_{h+1} = \varphi X_h, \quad h \geq 2.$$

Hence:

$$\alpha_X(1) = \mathrm{Corr}(\varphi X_1 + Z_2, X_1) = \varphi$$

$$\alpha_X(h) = \begin{cases} \varphi, & h = 1 \\ 0, & h \geq 2 \end{cases}$$

This is dual behavior to what function $\rho_X(h)$ does for MA(1) process.

EXAMPLE 7 (MA(1) & acf.)

Assume $Z_t \sim WN(0, \sigma^2)$

$$X_t = Z_t + \vartheta Z_{t-1}$$

$$\Rightarrow \quad \rho_X(h) = \begin{cases} \frac{\vartheta}{1+\vartheta^2}, & h = 1 \\ 0, & h \geq 2 \end{cases}$$

Always $\rho_X(0) = 1$, clearly.