Lecture-04

28/11/23 17/03/21

We will continue the concept of covariance from the last class

→ we came to Covariance because we wanted to analyze the problem of predicting one RV with the knowledge of the other RV.

$$e = E\left[|z_1 - \hat{z}_1|^2\right] \qquad \hat{z}_1 = a + b z_0, \quad a, b \in \mathbb{C}$$

our goal last lecture was to minimize the error 'e'

$$L(a,b) = E\left[|z_1 - \hat{z}_1|^2\right] = 0$$

$$\boxed{e_{min} = \sigma_{z_1}^2 \left(1 - \frac{|\gamma_{z_0 z_1}|^2}{\sigma_{z_0}^2 \sigma_{z_1}^2}\right) \geq 0}$$

minimum possible error

HW Task is to derive this expression.

e_{min} even though it is small should be Non-negative

$$\therefore |\gamma_{z_0 z_1}|^2 \leq \sigma_{z_0}^2 \sigma_{z_1}^2$$

Small covariance $\Rightarrow$ means $\underline{\gamma_{z_0 z_1} = 0}$

large covariance $\Rightarrow$ $\underline{\gamma_{z_0 z_1} = \sigma_{z_0}^2 \sigma_{z_1}^2}$

For small covariance case:

from the expression for a, b if $\gamma_{ab} = 0$

then $\underline{a = 0}$ & $\underline{b = m_{z_1}}$

& $\underline{e_{min} = \sigma_{z_1}^2}$

$\Rightarrow H(z_0) = a z_0 + b \Rightarrow 0 + m_{z_1}$

$$\boxed{H(z_0) = m_{z_1}}$$

The strange thing is it is the best possible choice for a, b

So, it is not possible to make a smaller error

But, interestingly the smaller error in this case of smaller covariance is achieved by discarding the input (z_0)

& Here in this case the best prediction ~~for~~ z_1 is it's average $\therefore$ $z_1 = m_{z_1}$ for $\gamma_{z_0 z_1} = 0$

In This case The Best we could do to predict z_1 is to say it's equal to it's Average without considering the other RV z_0.

i.e z_0, z_1 donot have information one on the other.

It is a Pathological (Stupid) case when z_0 carries no information on z_1

* This is not ~~exactly~~ True, entirely because it carriers no information That Can be exploited by Linear predictor.

on The contrary for large covariance case

$$\gamma_{z_0 z_1} = \sigma_{z_0}^2 \sigma_{z_1}^2$$

Then $e_{min} = 0$

Note: In this case there is perfect prediction.

$\therefore$ we can say Covariance is an Index of Linear predictability.

If we want to concentrate on the possibility of predicting and we want to remove the constraint that only predictor we can use is an Affine one

Q) What is the condition that allows us to say that two RV that are being considered can not be used for prediction one w.r.to another by any possible predictor?

Ans It turns out that the main concept that we have here is not Covariance but

it is **Statistical Independence**

To study it we will first describe Marginalization operation.

lets suppose we have two RVs

$$x_0, x_1 \in \mathbb{R} \qquad x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

$f_x \rightarrow$ Joint PDF

$\rightarrow$ Marginalization of PDF

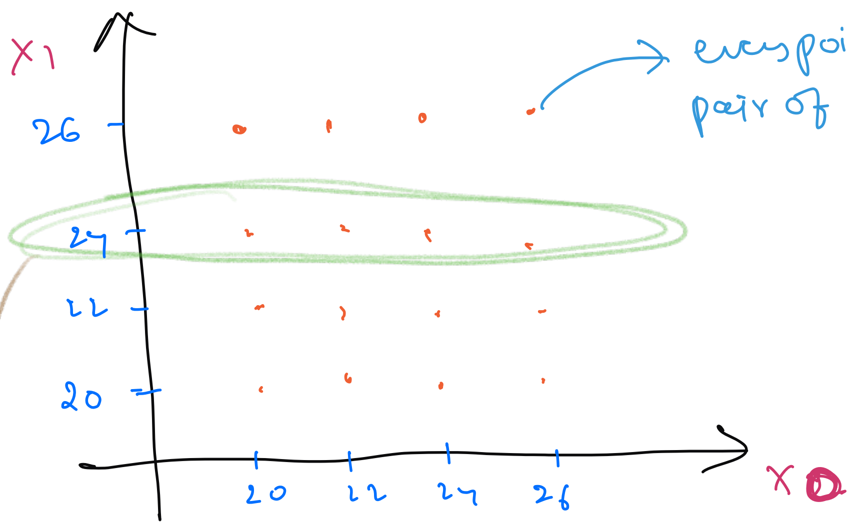
Marginal PDF w.r.to X_0

$$f_{X_0}(\xi_0) = \int_{-\infty}^{+\infty} f_X(\xi_0, \xi_1)\, d\xi_1$$

Marginal PDF w.r.to X_1

$$f_{X_1}(\xi_1) = \int_{-\infty}^{\infty} f_X(\xi_0, \xi_1)\, d\xi_0$$

ex: Two Rooms Example



Assume we could only distinguish from (20-27)
so this is not continuous distribution.

If we wanna know what is the probability of Room 1 @ 24 irrespective of the temperature of Room 0 i.e sum of all probabilities of Room 0 with Room 1 being at '22'

Definition: Two RV X_0 & X_1 are said to statistically independent only iff

$$f_X(\eta_0, \eta_1) = f_{X_0}(\eta_0)\, f_{X_1}(\eta_1)$$

their Joint probability is equal to the product of their Marginal probabilities.

Statistical Independence means zero covariance with the exception that it prevents predictability by any function 'H'

i.e not only by linear function but for any function.

→ we are trying to predict $\hat{X}_1$

$$\hat{X}_1 = H(X_0)$$

and we are wondering what is the function that minimizes the error function.

$$e = E\left[|X_1 - \hat{X}_1|^2\right]$$

$$= E\left[|X_1 - H(X_0)|^2\right]$$

$$= E\left[|X_1 - m_{X_1} + m_{X_1} - H(X_0) - m_{H(X_0)} + m_{H(X_0)}|^2\right]$$

Trick: Inside we subtract & add

$$m_{X_1} \quad \& \quad m_{H(X_0)}$$

$m_{H(X_0)}$: is the expectation of $H(X_0)$

$$m_{H(X_0)} = E\left[H(X_0)\right]$$

→ if we apply a function 'H' to a RV then we also get a random result.

$$\Rightarrow E\left[\left|\underbrace{(X_1 - m_{X_1})}_{A} + \underbrace{(m_{X_1} - m_{H(X_0)})}_{B} + \underbrace{(m_{H(X_0)} - H(X_0))}_{C}\right|^2\right]$$

$$\Rightarrow E\left[A^2 + B^2 + C^2 + 2AB + 2AC + 2BC\right]$$

$$\Rightarrow E\left[A^2\right] + E\left[B^2\right] + E\left[C^2\right] + 2E\underbrace{\begin{bmatrix} AB \\ AC \\ BC \end{bmatrix}}$$

single notation

they are equal to zero

we can show that

$$E[AB] = E[BC] = E[AC] = 0$$

$$E[AB] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{(x_1 - m_{x_1})}_{A} \underbrace{(m_{x_1} - m_{H(x_0)})}_{B} \underbrace{f_{\begin{bmatrix} x_0 \\ x_1 \end{bmatrix}}}_{\text{Joint Probability of } x_0, x_1} (\xi_0, \xi_1)\, d\xi_0\, d\xi_1$$

$\Rightarrow$ we know that Joint probability of x_0, x_1 can be equated as the product of Marginal probabilities of x_0, x_1

$$\boxed{\text{i.e } f_{\begin{bmatrix} x_0 \\ x_1 \end{bmatrix}} (\xi_0, \xi_1) = f(\xi_0) . f(\xi_1)}$$

for Statistically Independent

$$\Rightarrow \underbrace{(m_{x_1} - m_{H(x_0)})}_{\substack{\text{A Number} \\ \text{(constant)}}} \underbrace{\int_{-\infty}^{\infty} f_{x_0}(\xi_0)\, d\xi_0}_{=1} \boxed{\int_{-\infty}^{\infty} (\xi_1 - m_{x_1})\, f_{x_1}(\xi_1)\, d\xi_1}$$

Double Integral split as two Individual Integral

$\Rightarrow$

This integral is equal to Zero

$$\Rightarrow \underbrace{\int_{-\infty}^{\infty} \xi_1 f_{x_1}(\xi_1)\, d\xi_1}_{} - \boxed{\int_{-\infty}^{\infty} m_{x_1} f_{x_1}(\xi_1)\, d\xi_1}$$

This is the expectation of x_1

i.e m_{x_1}

This expectation of a constant which is equal to constant itself

$$\Rightarrow m_{x_1} - m_{x_1}$$

$$\Rightarrow 0$$

Similarly, $E[AC] = E[BC] = 0$

$$4 \quad e = \underbrace{E[(x_1 - m_{x_1})^2]}_{\sigma_{x_1}^2} + E[(m_{x_1} - m_{H(x_0)})^2] + E[(H(x_0) - m_{H(x_0)})^2]$$

error

(Variance of x_1)

$\sigma^2_{H(X_0)}$

expectation of constant = constant itself

$$= \left(m_{X_1} - m_{H(X_0)}\right)^2$$

$m_{X_1} = m_{H(X_0)}$ Then it is equal to zero

* → To minimize the Error

we have to reduce $\sigma^2_{H(X_0)}$, $\left(m_{X_1} - m_{H(X_0)}\right)^2$

To reduce a positive quantity we have to make it zero.

i.e we have to make the variance of $H(X_0)$

$$\sigma^2_{H(X_0)} = 0$$

So, Here we use Chebychev Inequality

i.e $Pr\left\{\left|H(X_0) - m_{X_1}\right| > \varepsilon\right\} \leq \dfrac{\sigma^2_{H(X_0)}}{\varepsilon^2}$

$\therefore$ we have to make $\sigma^2_{H(X_0)} = 0$

$$\Rightarrow \Pr\{|H(X_0) - m_{X_1}| > \varepsilon\} = 0$$

Since $H(X_0)$ cannot deviate from m_{X_1}

$$\therefore \quad H(X_0) = m_{X_1}$$

This happens independently of $H(X_0)$

we are exactly having the same condition we had during the calculation of Zero Covariance.

i.e Zero covariance we put for Linear Predictor

Note; When we have Statistical Independence The input (X_0) is Useless because the best prediction we could do is to predict $X_1 = m_{X_1}$ & this is the strategy that minimizes the Error

$\rightarrow$ So, Zero covariance means I cannot predict anything from the linear Predictor.

Statistical Independence means I cannot predict anything from any predictor.

→ Linear Predictor using many Data

$$X = \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \end{bmatrix} \in \mathbb{C}^n \qquad w \in \mathbb{C}$$

Data ↙

$$H : \mathbb{C}^n \rightarrow \mathbb{C}$$

all Data is arranged in a vector form, and we assume it to be complex vector just to make it General.

W is an Unknown vector, that has to predicted

And our predictor function $H : \mathbb{C}^n \rightarrow \mathbb{C}$

$$\hat{w} = H(X) = a^{+} X$$

where $a \in \mathbb{C}^n$ $\quad a = \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \end{bmatrix}$

$$\Rightarrow \boxed{H(X) = \sum_{j=0}^{n-1} a_j^{*} x_j}$$

Vector scalar product

we can also add an offset b

$$\hat{x} = \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \\ 1 \end{bmatrix} \qquad \hat{a} = \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \\ b \end{bmatrix}$$

$1 = x^0$ $\qquad$ $b \rightarrow$ offset

$$\boxed{\hat{w} = \hat{a}^t \hat{x} = a^t x + b}$$

→ Now we can proceed to calculate the ERROR

$$e = E\left[|w - \hat{w}|^2\right] = E\left[|w - a^t x|^2\right]$$

$$= E\left[(w - a^t x)(w - a^t x)^*\right]$$

Comments by Prof:

- What we will do now, is a derivation that gives us two kinds of Results

- The first result is what we land in Covariance calc. if one variable is known and other variable to predict and that will gives us the coefficients, that is the optimal value of the coefficient vector 'a'

- While we are doing so, we will go through some passages that will have significance for us and we will save that passages as derivations for later, to comment from a different point of view.
- So our derivations would be doubled, I will make our calculations to arrive at optimal 'a' and during this derivation, we will do something and save them for later.
- Later, we will comment on them, the so called orthogonality principle that gives us a Geometric Interpretation of task of predicting linear by Things.

→ Let's begin with the Derivation for Minimum error!

$$e = E\left[\,|w-\hat{w}|^2\,\right] = E\left[\,|w-a^{\dagger}x|^2\,\right]$$

Average Quadratic Error!

$$= E\left[(w-a^{\dagger}x)\,\underbrace{(w-a^{\dagger}x)^*}_{\text{complex conjugation}}\right]$$

* We want to minimize the Error so we have to find an optimum 'a'

But, Thanks to The Branwood trick

So, we will take a psuedo gradient w.r.to a^*

$$\tilde{\nabla}_{a^*} e = 0$$

Interestingly in $e = \left[(w - a^\dagger x) \, (w - a^\dagger x)^* \right]$

Only This term has a^*

$a^\dagger$ = 'a' conjugate Transpose

This does not result in having a^* because it is the conjugation of $a^\dagger$ ('a' conjugate Transpose)

$\therefore$ when you derive 'e' w.r.to a^* it results in

$$\tilde{\nabla}_{a^*} e = 0$$

$$\Rightarrow \frac{d}{da^*} E\left[(w - a^\dagger x) \, (w - a^\dagger x)^* \right]$$

$$\Rightarrow E\left[0 - x (w - a^\dagger x)^* \right] = 0$$

When we make a psuedo derivative w.r.to something conjugated we have to consider the non conjugated version as fixed.

$$\Rightarrow E\left[-x\,\underbrace{(w-a^\dagger x)^*}_{\varepsilon^*}\right]=0$$

Save for later

$$E[x\varepsilon^*]=0$$

$$\varepsilon = w-\hat{w}$$

$$e = E[|\varepsilon|^2]$$

1

$$E[x(w-a^\dagger x)^\dagger]=0$$

Here we changed a conjugate into a conjugate transpose.

$$\Rightarrow E[x(w-a^\dagger x)^\dagger]=0$$

$$\Rightarrow E[x(w^\dagger - x^\dagger a)]=0$$

$$\Rightarrow \underbrace{E[xw^\dagger]}_{b} - \underbrace{E[xx^\dagger]}_{R_x}\,a=0$$

$$\Rightarrow \quad b \in \mathbb{C}^n \qquad R_x \in \mathbb{C}^{n\times n}$$

$$\underbrace{R_x = E\left[\begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \end{bmatrix}\begin{bmatrix} x_0^* & \cdots & x_{n-1}^* \end{bmatrix}\right]}_{\text{correlation Matrix}}$$

$$\Rightarrow \quad b - R_x a = 0$$

$$\Rightarrow \quad \underbrace{\boxed{a = R_x^{-1} b}}_{\text{Optimal Parameters}}$$

→ In principle we have solved our First problem i.e finding 'a' i.e optimal set of coefficients allowing us to predict '$\hat{w}$'

From Now on we will do calculation in Two colums Left columns will be Analogous to what we did earlier in the case of covariance, Right column will be The provide with some properties of Hermetian Matrix R_x

Left column | Right column

$e_{min} = ?$

Minimal Quadratic error

To compute it we put the formula for optimal 'a' i.e $a = R_x^{-1} b$

$$\Rightarrow e_{min} = E\left[|w - a^T x|^2 \right]$$

$$= E\left[\left| w - (R_x^{-1} b)^T x \right|^2 \right]$$

$$= E\left[\left(w - (R_x^{-1} b)^T x\right) \left(\left(w - (R_x^{-1} b)^T x\right)^{*} \right) \right]$$

$+$

$$\Rightarrow E\left[\left(w - (R_x^{-1} b) x\right) \left(w^{*} - x^T (R_x^{-1} b) \right) \right]$$

$$\Rightarrow E[w w^{*}] + E\left[b^T (R_x^{-1})^T x x^T R_x^{-1} b \right] - E\left[b^T (R_x^{-1})^T x w^{*} \right]$$

- R_x is Hermitian

i.e $R_x^{\dagger} = R_x$

$\dagger$ = conjugate & Transpose

- R_x is Positive Semidefinite

$\forall u \in \mathbb{C}^n$

$$\boxed{u^{\dagger} R_x u \geq 0}$$

$u^{\dagger} E[x x^{\dagger}] u$

$$\Rightarrow E[\underbrace{u^{\dagger} x}_{y^{*}} \underbrace{x^{\dagger} u}_{y}]$$

$$\Rightarrow E[y^{*} y]$$

$$\Rightarrow E[|y|^2] \geq 0$$

$\because$ expectation of Square Modulo is always Greater Than Zero

$$\boxed{- E[w x^{\dagger} R_x^{-1} b]}$$

$$\Rightarrow E[ww^*]$$

$$+ \; b^{\dagger} R_x^{-1} \underbrace{E[x x^{\dagger}]}_{= R_x} R_x^{-1} b$$

$$- \; b^{\dagger} R_x^{-1} \underbrace{E[x w^*]}_{b}$$

$$- \; \underbrace{E[w x^{\dagger}]}_{b^{\dagger}} R_x^{-1} b$$

$$\Rightarrow E[ww^*] + b^{\dagger} R_x^{-1} \cancel{R_x R_x^{-1}} b - b^{\dagger} R_x^{-1} b - b^{\dagger} R_x^{-1} b$$

$$\Rightarrow E[ww^*] + \cancel{b^{\dagger} R_x^{-1} b} - \cancel{b^{\dagger} R_x^{-1} b} - b^{\dagger} R_x^{-1} b$$

$$\Rightarrow E[ww^*] - b^{\dagger} R_x^{-1} b$$

R_x^{-1} is also Hermetian

$$R_x R_x^{-1} = I \quad \because R_x = R_x^{\dagger}$$

$$\Rightarrow R_x^{\dagger} R_x^{-1} = I$$

$\Rightarrow$ Multiply by $(R_x^{-1})^{\dagger}$ on both sides

$$\Rightarrow (R_x^{-1})^{\dagger} R_x^{\dagger} R_x^{-1} = (R_x^{-1})^{\dagger}$$

$$\Rightarrow (R_x R_x^{-1})^{\dagger} R_x^{-1} = (R_x^{-1})^{\dagger}$$

$$\Rightarrow \underbrace{(I)^{\dagger} R_x^{-1} = (R_x^{-1})^{\dagger}}$$

$\therefore$ R_x^{-1} is also Hermetian

$\therefore$ Final expression for e_{min}

$$\boxed{e_{min} = E[ww^*] - b^+ R_x^{-1} b}$$

from save for later

$$e_{min} = E[ww^*] - b^+ R_x^{-1} E[xw^*]$$

$$= E[ww^*] - E[b^+ R_x^{-1} x w^*]$$

$$= E[(w - \underbrace{b^+ R_x^{-1} x}_{\hat{w}}) w^*]$$

$$= E[\varepsilon w^*]$$

2

$$\varepsilon = w - \hat{w}$$

$$\boxed{e_{min} = E[w \varepsilon^*]}$$

$\therefore$ e_{min} is constant and ≥ 0

we can conjugate e_{min}

From save for later

$$\boxed{E[x\varepsilon^*]=0 \quad E[w\varepsilon^*]=e_{min}} \qquad \varepsilon = w-\hat{w}$$

condition for optimality

Minimum possible error by linear prediction

- We will use them to give a Geometric Interpretation to the process of making a linear prediction.

- The above conditions represent Orthogonality principle

$$\boxed{E[x\varepsilon^*]=0}, \quad \boxed{E[w\varepsilon^*]=e_{min}}$$

We can make Geometric Interpretation by interpreting RV's as Vectors

* In the Lecture-1 we generalized the concept of vectors from just Arrows in space to Functions which satisfy certain Axioms.

Now we will generalize it to Random variables.

→ vector Space of RV's

α, β random variables

distributes w.r.to sum

- $3\alpha + 2\beta$ is Random variable itself

- $\underbrace{\langle \alpha, \beta \rangle}_{\text{Scalar product}} = E[\alpha^* \beta]$ satisfies axioms of Scalar product

$\langle \alpha, \beta \rangle = 0 = E[\alpha^* \beta] \Rightarrow \alpha \perp \beta$ (perpendicular) (orthogonal)

once we have Vector space & Scalar product we have a whole Euclidean Geometry on my set of vectors.

→ We now can define Norm i.e length of the RV α

$$\|\alpha\| = \sqrt{\langle \alpha, \alpha \rangle} = \sqrt{E[\alpha^* \alpha]} = \sqrt{E[|\alpha|^2]}$$

$$e = E[|\varepsilon|^2] \Rightarrow \boxed{e = \|\varepsilon\|^2}$$

Geometric Interpretation of Linear prediction

ex:

$n = 2 \quad X = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$

'w' predicted variable

$\hat{w} = a^* X \qquad a = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$

$\hat{w}$ actual prediction

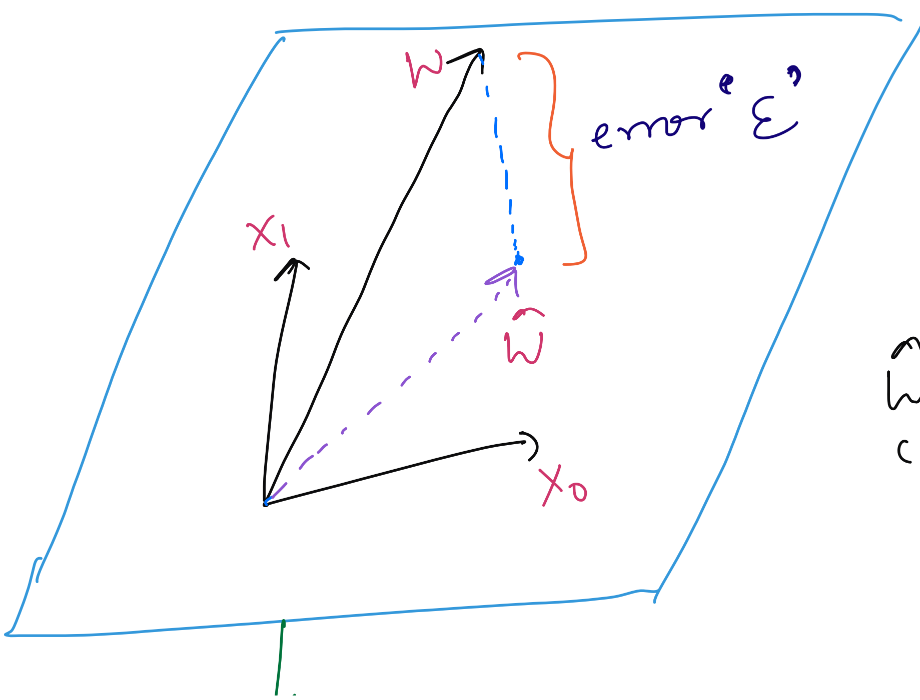
$\hat{w} = [a_0 \ a_1]^* \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$

$\varepsilon = w - \hat{w}$ (error)

$\Rightarrow [a_0^* \ \ a_1^*] \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$

$\hat{w} \Rightarrow a_0^* x_0 + a_1^* x_1$

→



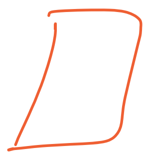
$\hat{w}$ = linear combination of x_0, x_1

This Plane is Set of all possible Linear predictions.

∴ it can be described by the Linear combinations of x_0, x_1

→ Conditions for optimality

$\varepsilon = w - \hat{w}$

$\varepsilon \perp$ 

$\varepsilon \perp x_0$ & $\varepsilon \perp x_0$

finding the $\hat{w}$ which is close to w since all our predictions $\hat{w}$ must stay on the plane, from the picture it's the error which is the distance from Head of w to the Head of $\hat{w}$.

This distance should be as Minimum as possible & in Eucledian Geometry and we are doing Eucledian Geometry this means we are doing Orthogonal projection of w onto the plane

once you do that we can identify the projection of w i.e $\hat{w}$ as the best possible prediction.

$E[x\varepsilon^*]$ → This is the condition for optimality which we can interpret as vector equality.

$$\Rightarrow E[X\varepsilon^*] = E\left[\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \varepsilon^*\right]$$

$$\Rightarrow \underbrace{\begin{bmatrix} E[x_0\varepsilon^*] \\ E[x_1\varepsilon^*] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\text{Vector Equality}}$$

(I)

Note:

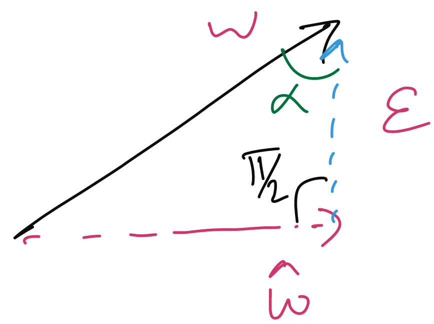
→ $\boxed{\text{To be optimal a Linear Predictor must commit Errors orthogonal to the Data.}}$

→ Geometric interpretation of second equality

$$e_{min} = E\left[|\varepsilon|^2\right] = E\left[\varepsilon^*\varepsilon\right]$$

$$\Rightarrow \underbrace{\langle \varepsilon, \varepsilon \rangle = \|\varepsilon\|^2}_{\text{always happens in space that has a scalar product}}$$

let's focus on the figure Δ^{le} above with orthogonal projection.



→ The orthogonality principle states that

$$e_{min} = E\left[\varepsilon^* w\right] = \langle \varepsilon , w \rangle$$

Scalar product

$$= \underbrace{\|\varepsilon\| \, \|w\| \cos(\alpha)}_{\|\varepsilon\|}$$

Note:

$$= \|\varepsilon\|^2$$

> The Minimum Average Squared ERROR that a linear predictor can make is the projection of the instantaneous error onto the variable to be predicted (or) vice versa

→ Orthogonality Principle can be applied in the other tasks of linear prediction.

In particular, when we will use it in deriving linear predictors that do not predict a value from a vector space of Data, But Try to predict a process i.e future of a process starting from its past

→ Characteristic function of a RV :

$$X \in \mathbb{R} \qquad \psi_X(\omega)$$
(RV)

→ Generally, for a RV 'X' we usually derive its behaviour in PDF.

There is also an another way in terms of Characteristic function which is indicated as above.

The characteristic function has three definitions that are all equivalent.

variable to define characteristic function.

$$\psi_X(\omega) \overset{①}{=} E\left[e^{2\pi i \omega X}\right]$$

it is the Expectation of Complex Exponential defined above.

② $\mathcal{F}^{-1}[f_X](\omega) \Rightarrow$ Inverse F-T of PDF of 'X' in ω

③ $\mathcal{F}[f_X](-\omega) \Rightarrow$ F.T of PDF of 'X' in $-\omega$

- Therefore once we know the PDF we can compute the characteristic function and vice versa

properties of ψ_X * They are very useful in Gaussian case

1) $\psi_X(-\omega) = \psi_X^*(\omega)$

2) $\psi_X(0) = 1$

3) $y = ax + b$ $\quad \boxed{\psi_Y(\omega) = e^{2\pi i \omega b} \cdot \psi_X(a\omega)}$

4) $\dfrac{\partial^p}{\partial \omega^p} \psi_X(\omega) = \dfrac{\partial^p}{\partial \omega^p} E\left[e^{2\pi i \omega X}\right]$

$$= E\left[\frac{\partial^p}{\partial\omega^p} e^{2\pi i\omega x}\right]$$

$$= (2\pi i)^p\, E\left[x^p e^{2\pi i\omega x}\right]$$

for $\omega = 0$

$$\boxed{\underbrace{\frac{\partial^p}{\partial\omega^p}\varphi_X(0)}_{\text{Moment Generating Function}} = (2\pi i)^p \underbrace{m_X{}^p}_{\text{noncentral order of 'p'}}}$$

End comments:

→ Now that we have exhausted the topics that don't change in time - I want to extend the Randomness property to Quantities That vary in time (waveforms)

i.e we want to find a way of Modelling Randomness to Signals that are quantities

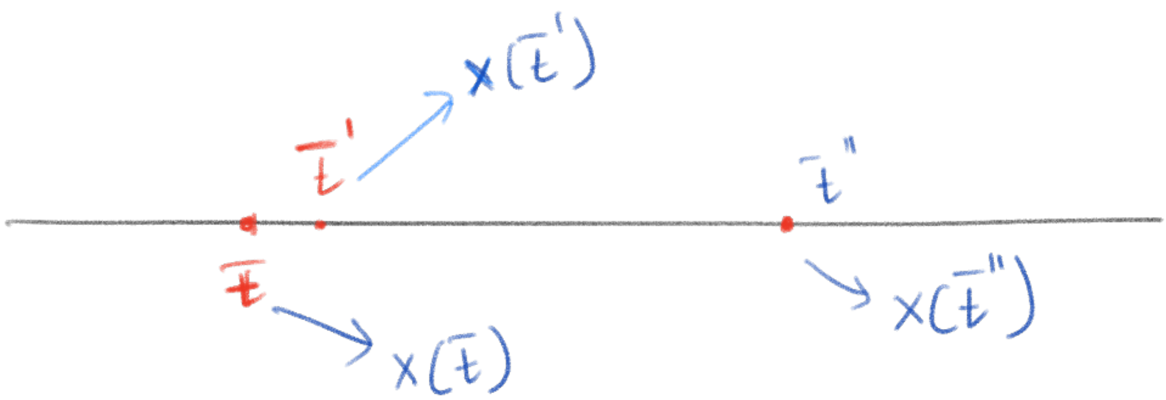
That change in Time to carry information.

* Stochastic Process (Random process)

The Mathematical entity that embodies this mechanism of Modelling Randomness of a Single That vary in time and also vary Randomly is called Stochastic process (or) Random Process

Stochastic process can be seen in Two ways we will begin with the most elementary way and Then we will realize That there is a more Sophisticated and effective way to look at stochastic processes.

Idea 1: we have a Time axis and for every time instant there is a Random variable



That gives us the value of the wave form at respective time $\bar{t}$

or @ t' $\longrightarrow$ $x(t')$

@ t'' $\longrightarrow$ $x(t'')$

Random variable @ specific time interval

Definition 1: For each 't' gives us a random variable $X(t)$

∴ Since we have many RV one for every possible time instant.

Q) How to characterize the stochastic processes?

Ans In a single Random Variable case we can know its PDF we can know everything

i.e Characteristic function.

∴ In case of stochastic process things get complicated but not unmanageable.

Because what we have to convey is that, we must be able to give me something for every choice of the Time Instance.

So $\forall\ m$ $\forall$ $\underbrace{t_0 < t_1 < t_2 < \cdots t_{m-1}}_{\text{Choices of Time Instances}}$

→ Here, we want to know Q) What is the Joint probability Density of the RVs ?

$$\boxed{\text{i.e } f_{X(t_0),X(t_1)\cdots X(t_{m-1})}(\zeta_0,\zeta_1,\cdots\zeta_{m-1})}$$

$$\boxed{f : \mathbb{R}^m \longrightarrow \mathbb{R}^+}$$

Our Joint PDF would have 'm' inputs & one +ve o/p.

→ By this procedure we have to do the above thing for every choice of 'm' & this means that to know everything about stochastic process we have to know about every time instant if $m \longrightarrow \infty$ & This is not feasible.

JPDF is only used for $m=1$ & $m=2$

$m=1$ first order characterization $f_{X(t)}(\zeta)$

$m=2$ Second order characterization $f_{X(t_0)X(t_1)}(\zeta_0,\zeta_1)$

* Correlation / covariance characteristic

$$\forall m \quad \forall \; t_0 < t_1 < \cdots < t_{m-1}$$

(i) Correlation Function $\Rightarrow$ $C_X(t_0, t_1, \cdots t_{m-1})$

If complex

$$C_X(t_0, t_1, \cdots t_{m-1}) = E\left[X^*(t_0)\, X(t_1)\, X(t_2) \cdots \cdots X(t_{m-1}) \right]$$

(ii) Covariance function : $\rightarrow$ $K_X(t_0, t_1, \cdots t_{m-1})$

$$= E\left[\left(X(t_0) - m_{X(t_0)}\right)^* \left(X(t_1) - m_{X(t_1)}\right) \left(\_\_\_\right) \cdots\cdots \left(X(t_{m-1}) - m_{X(t_{m-1})}\right) \right]$$

$m_{X(t_0)} = E[X(t_0)]$

$* \rightarrow$ If complex

* The Conjugation is alternative. i.e every

other 'X' is conjugated.

→ Integral representation of Correlation

$$\boxed{C_x(t_0, t_1, \dots t_{m-1}) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{m-\text{times}} \left(\zeta_0^* \zeta_1 \zeta_2^* \cdots \zeta_{m-1}\right) \underbrace{f_{x(t_0)\dots x(t_{m-1})}}_{\text{Joint PDF}} (\zeta_0, \zeta_1, \dots \zeta_{m-1}) \, d\zeta_0 \, d\zeta_1 \dots \dots d\zeta_{m-1}}$$

This is strict equivalent to Covariance, if you know the Joint PDF then you can compute C_x but the Reverse is not True.

→ Since, JPDF characterisation is a finer characterisation than covariance. Still giving a function for every possible value 'm' from (0 to m-1) is an excessive task

∴ Correlation is used mostly for 2nd order at most 3rd & 4th

→ Second order correlation (Distribution of Energy)

$m=2$ $\qquad C_X(t_0,t_1)$ is Positive SemiDefinite

In most case 2^{nd} order correlation is sufficient because it gives us information about Energy distribution in Time in our waveform

properties of 2^{nd} order correlation:

$C_X(t_0,t_1)$ can be seen as a Generalization of a Matrix for Quadratic form.

Proof:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{C_X(t_0,t_1)}_{A_{jk}} \underbrace{\phi(t_0)}_{U_j} \underbrace{\phi(t_1)}_{U_k}\, dt_0\, dt_1$$

$U^T A U$ → Quadratic form

$$= \sum_j \sum_k A_{jk} U_J U_K$$

$$\Rightarrow \iint E[x(t_0)\,x(t_1)]\,\phi(t_0)\,\phi(t_1)\,dt_0\,dt_1$$

$$\Rightarrow E\Big[\underbrace{\int x(t_0)\,\phi(t_0)\,dt_0 \int x(t_1)\,\phi(t_1)\,dt_1}\Big]$$

double integral can be seperated into Two single Integrals. & They are the same

$$\Rightarrow E\left[\left(\int x(t)\,\phi(t)\,dt\right)^2\right] \geq 0$$

$\therefore$ Quadratic form is alway non-negative

* Correlation matrices are +ve semi-definite

They are also Hermetian

$$C_X(\underbrace{t_1, t_0}_{\text{exchanged}}) = E[x(t_1)\,x^*(t_0)]$$

$$= C_X^*(t_0, t_1)$$