MIDA I

Noé DEBROIS

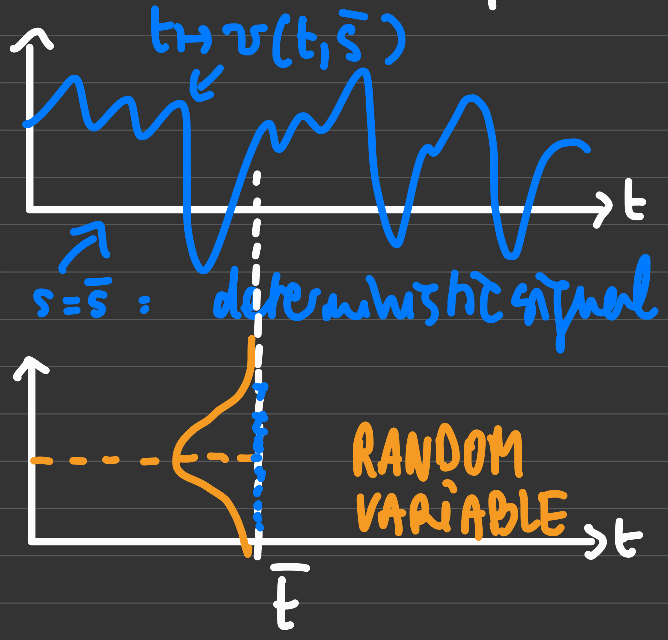
Politecnico di Milano

2023/2024

Stochastic process: (SP) it's an $\underline{\infty \text{ sequence}}$ of random variables, all defined on the same probabilistic space.

$$\ldots\ v(1,s),\ v(2,s),\ \ldots$$

time idx. — random experiment realization.



Describing a S.P by its mean & covariance function is sufficient for our applications.

Mean value of a SP:

$$m(t) = \mathbb{E}\left[v(t,s)\right] = \int_{\mathbb{R}} v(t,s)\,\mathrm{pdf}(s)\,ds.$$

Covariance fct of a SP:

$$r(t_1,t_2) = \mathbb{E}\left[\left(v(t_1,s) - m(t_1)\right)\left(v(t_2,s) - m(t_2)\right)\right].$$

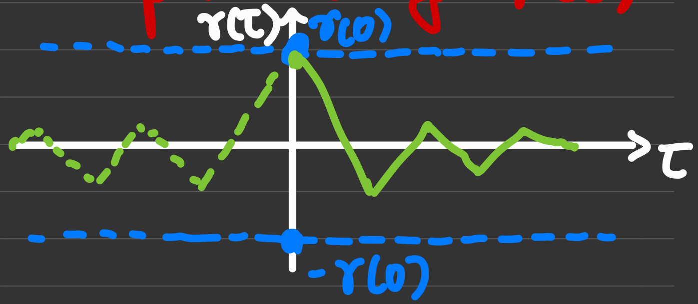
(If $t_1 = t_2$: variance of the SP).

Stationary SP (SSP): It's a SP with:

1. $\forall t,\ m(t) = m$;
2. $\gamma(t_1, t_2)$ depends on $\tau = t_2 - t_1$ only.

In this case, γ has some nice properties.

Properties of $\gamma(\tau)$:



1. $\gamma(0) = \mathbb{E}[(v(t,s) - m)^2] \geq 0$ POSITIVITY.
2. $|\gamma(\tau)| \leq \gamma(0),\ \forall \tau$ NON INCREASING.
3. $\forall \tau,\ \gamma(\tau) = \gamma(-\tau)$ EVEN.

⚠ Covariance fct $\neq \underbrace{\mathbb{E}[v(t_1,s)\,v(t_2,s)]}_{\text{CORRELATION FCT.}}$ if $m \neq 0$

WHITE NOISE: (WN) it's a SSP with mean μ, variance λ^2 $(e(t) \sim WN(\mu, \lambda^2))$ and such that:

1. $\mathbb{E}[e(t)] = \mu$;

② $\mathbb{E}\left[(e(t)-\mu)^2\right] = \lambda^2$;

Correlation between ANY PAIR of samples $= 0$.

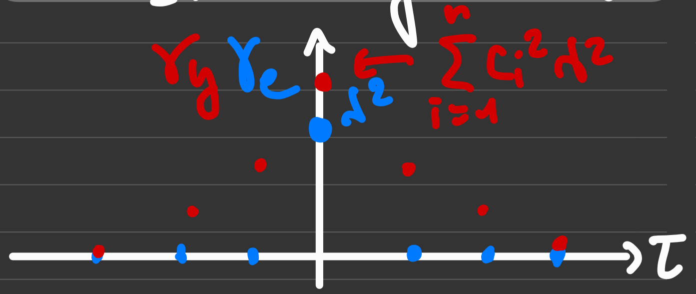
③ $\gamma_e(\tau) = \mathbb{E}\left[(e(t)-\mu)(e(t+\tau)-\mu)\right] = 0$, $\forall t, \forall \tau \neq 0$.

MOVING AVERAGES

M.A process of order n: $(MA(n))$ if $e(t) \sim WN(0,\lambda^2)$, it is the process obtained as:

[Adapt $\{c_i\}_i$ to give ± weight to the old/recent steps.]

$$y(t) = c_0 e(t) + c_1 e(t-1) + \dots + c_n e(t-n).$$



Stationarity of MA(n): MA(n) is still stationary

MA(∞): generalization of MA(n):

$$y(t) = \sum_{i=0}^{+\infty} c_i \, e(t-i), \quad e(t) \sim WN(0,\lambda^2)$$

⚠ with assumption: $\sum_{i=0}^{+\infty} c_i^2 < +\infty$.

Properties of MA(∞): Mean : $m(y(t)) = 0$;

Variance : $\gamma_y(t,t) = \sum_{i=0}^{+\infty} c_i^2 \cdot \lambda^2 < +\infty$.

Covariance : $\gamma_y(t,t-\tau) = \sum_{i=0}^{+\infty} c_i c_{i+\tau} \cdot \lambda^2$.

MA(∞) is a SSP

↳ It can be used to model almost all the SSP of this course.

AUTOREGRESSIVE MODELS

Allow us to have $\gamma(\tau) \neq 0$ for all τ, with a finite set of coeff.

↓

More "handy" than MA(∞)

AR(1) : $y(t) = a y(t-1) + e(t)$

$= a(a y(t-2) + e(t-1)) + e(t)$

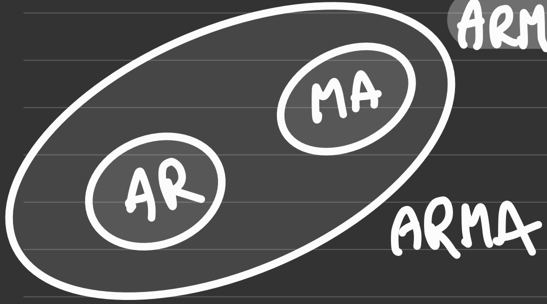
$= \dots$

AR(m): Let $e(t) \sim WN(0, \lambda^2)$, consider:

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + \dots + a_m y(t-m) + e(t)$$

Prop : AR(1) $\equiv$ MA(∞) with $c_0 = 1, c_1 = a, \dots$.

But the modeling power of MA(∞) is still higher than that of a AR(1), as the coefficients are severely constrained.



ARMA(m,n): Let $e(t) \sim WN(0,\lambda^2)$ and consider:

$$y(t) = a_1 y(t-1) + \dots + a_m y(t-m) + \quad \leftarrow AR(m)$$

$$c_0 e(t) + c_1 e(t-1) + \dots + c_n e(t-n) \quad \leftarrow MA(n).$$

ARMAX(m,n,p,k):

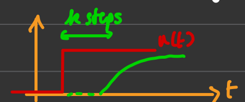
$$y(t) = a_1 y(t-1) + \dots + a_m y(t-m) + \quad \leftarrow AR(m)$$

$$c_0 e(t) + \dots + c_n e(t-n) + \quad \leftarrow MA(n)$$

$$b_0 u(t-k) + \dots + b_p u(t-k-p) \quad \leftarrow X(k,p)$$

"Exo-genous".

Real input!



k: pure input/output (i/o) delay;

p: order of the exogenous part.

Non-linear ARMAX:

$$y(t) = f\big(y(t-1), \dots, y(t-m),$$
$$e(t), \dots, e(t-n),$$
$$u(t-k), \dots, u(t-k-p)\big)$$

THE OPERATORIAL REPRESENTATION

with f that can be any nonlinear, parametric function:

- polynomials,
- splines,
- wavelets,
- neural networks ...

And all where you have an AR part.

We need + tools to study stationarity of AR models

BACKWARD-SHIFT OPERATOR: z^{-1}, defined as: $z^{-1} \cdot x(t) = x(t-1)$.

FWD-SHIFT OPERATOR: $z \cdot x(t) = x(t+1)$.

PROPERTIES:

LINEARITY: $z^{-1} \cdot (a x(t) + b y(t)) = a x(t-1) + b y(t-1)$.

RECURSIVE: $z^{-1}(z^{-1}(z^{-1}(x(t)))) = z^{-3} \cdot x(t)$
$= x(t-3)$.

LINEAR COMPOSITION: $(a z^{-1} + b z)\, x(t) = a x(t-1) + b x(t+1)$

DISCRETE TIME TRANSFERT FCT:

$W(z)$

Given an ARMA(m,n):

$$y(t) = a_1 y(t-1) + \dots + c_0 e(t) + \dots + c_n e(t-n)$$

$$= a_1 z^{-1} y(t) + \dots + c_0 e(t) + \dots + c_n z^{-n} e(t)$$

i.e:

$$\left(1 - a_1 z^{-1} - \dots - a_m z^{-m}\right) y(t) = \left(c_0 + c_1 z^{-1} + \dots + c_n z^{-n}\right) e(t)$$

i.e: $$y(t) = \frac{c_0 + c_1 z^{-1} + \dots + c_n z^{-n}}{1 - a_1 z^{-1} - \dots - a_m z^{-m}} e(t) = \frac{C(z)}{A(z)} e(t)$$

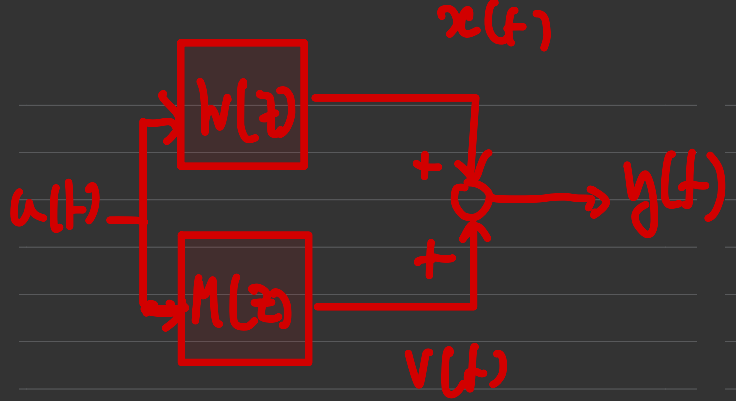
$$y(t) = W(z) e(t)$$

OPERATIONS WITH TF:

$$\xrightarrow{u(t)} \boxed{W(z)} \xrightarrow{x(t)} \boxed{M(z)} \xrightarrow{y(t)}$$

SERIES INTERCONNECTION

$$y(t) = \underbrace{M(z) W(z)}_{\text{Equivalent TF.}} u(t)$$



$$y(t) = \underbrace{\big(W(z) + M(z)\big)}_{\text{Equivalent TF}} u(t)$$

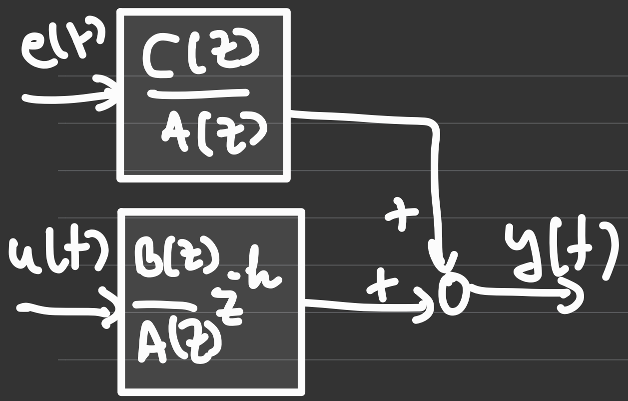
TF FOR ARMAX :

$$\begin{aligned} y(t) = {} & a_1 y(t-1) + \dots + a_m y(t-m) + \\ & c_0 e(t) + \dots + c_n e(t-n) + \\ & b_0 u(t-k) + \dots + b_p u(t-k-p) \end{aligned}$$

So using the same trick:

$$\underbrace{\left(1 - a_1 z^{-1} - \dots - a_m z^{-m}\right)}_{A(z)} y(t) =$$

$$\underbrace{\left(c_0 + c_1 z^{-1} + \dots + c_n z^{-n}\right)}_{C(z)} e(t) + \underbrace{\left(b_0 + b_1 z^{-1} + \dots\right) z^{-k}}_{B(z)} u(t)$$



$$S_0: \; y(t) = \frac{B(z)}{A(z)} z^{-k} u(t) + \frac{C(z)}{A(z)} e(t),$$

$$e(t) \sim WN(0, \lambda^2)$$

↳ $y(t)$ depends on the following, in ARMAX models:

- 2 signals: $u(t)$ [REAL] and $e(t)$ [FICTITIOUS];
- 3 polynomials: $A(z)$ [AR], $B(z)$ [MA], $B(z)$ [X].

ZEROS OF $W(z)$: For $W(z) = \dfrac{C(z)}{A(z)}$, the zeros of $W(z)$ are the values of z s.t. $W(z) = 0$.

POLES OF $W(z)$: The poles of $W(z)$ are the values of z such that $W(z)^{-1} = 0$.

THEOREM: A linear digital filter w/ TF $W(z)$ is <u>asymptotically stable</u> IFF all its poles are strictly inside the unit circle.
→ If zeros are inside: the filter is "minimum phase".

COROLLARY FOR MA(n)/AR(m):
- MA(n) has:
  - n non trivial zeros;
  - n poles, all lying at the origin.

  ⟺ all MA(n) processes are asymptotically stable.

- AR(m) has:
  - m zeros all lying at the origin;
  - m non-trivial poles.

  ⟺ All AR(m) processes are minimum phase.

(⚠ for AR to be stable: $|a| < 1$ ... (cf. econometrics)).

THEOREM: $\xrightarrow{e(t)} \boxed{W(z)} \xrightarrow{y(t)}$ the steady-state output

$y(t)$ is STATIONARY IFF:

- $e(t)$ is stationary;
- $W(z)$ is asymptotically stable.

Mean & Covariance function of ARMA models

AR(1) & MA(∞) are "equivalent" (thanks to infinite division). It's not easy to compute mean & cov. fct based on an ∞ amount of parameters $e_0, e_1, \dots$.

↳ we need to derive another computation method based on the def. of AR model.

MEAN, VAR, COV ... AR(1):

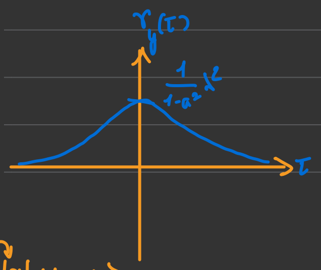
$$y(t) = a y(t-1) + e(t)$$

> Proof of $\mathbb{E}[y(t)e(t-1)] = \mathbb{E}[y(t)e(t-2)] = \dots = 0$:
> Equivalence between AR(1) & MA(∞): remember that
> $y(t) = e(t) + a e(t-1) + a^2 e(t-2) \dots$ so it's clear. $\square$

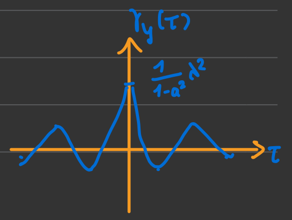
- $m_y = a \underbrace{\mathbb{E}[y(t-1)]}_{=m_y} + \underbrace{\mathbb{E}[e(t)]}_{=0}$ i.e. $m_y = \frac{1}{1-a} m_e = 0$.
- VAR: $\gamma_y(0) = \mathbb{E}\left[(y(t) - \cancel{m_y})^2\right] = \mathbb{E}[y(t)^2] = \frac{1}{1-a^2} \lambda^2$. same technique as for m_e
- COVAR:

<u>Yule Walker Equations</u>:
$$\begin{cases} \gamma_y(0) = \frac{1}{1-a^2}\lambda^2 \\ \gamma_y(\tau) = a\,\gamma_y(\tau-1), \ |\tau| \geq 1 \end{cases}$$
(for AR(1) proc.)

(compute $\gamma_y(0)$ & then $\gamma_y(1), \dots$ you will naturally find a recursive equation).



SSP
$|a|<1, a>0$



$|a|<1, a<0$

SAME FOR ARMA(m,n):

$y(t) = a_1 y(t-1) + \dots + a_m y(t-m) + c_0 e(t) + \dots + c_n e(t-n)$

(assume $a_1, \dots, a_m$ are such that $y(t)$ is SSP)

- MEAN: $m_y = 0$

- $\gamma_y(0) = \mathbb{E}[y(t)^2] = a_1^2 \mathbb{E}[y(t-1)^2] + \dots + c_n^2 \mathbb{E}[e(t-n)^2]$

because:

$\gamma_y(0) = \mathbb{E}[y(t)^2] = \mathbb{E}[(a_1 y(t-1) + \dots + a_m y(t-m) + c_0 e(t) + \dots + c_n e(t-n))^2] \dots$

$+ 2a_1a_2 \mathbb{E}[y(t-1)y(t-2)] + \dots$ ← $\gamma_y(1)$ and other γ_y's.

$+ 2a_1c_0 \mathbb{E}[y(t-1)e(t)] + \dots$ ← Products of AR and MA parts ...

$+ \cancel{2c_0c_1 \mathbb{E}[e(t)e(t-1)]} + \dots$

$= 0$

$\gamma_y(1) = a_1 \gamma_y(0) + \dots$

$\vdots$

$\gamma_y(m-1) = a_1 \gamma_y(m-2) + \dots$

m recursive eq. for the initialization of Yule-Walker Equations for ARMA(m,n).

What about proc. with non 0 mean?

If $e(t) \sim WN(\mu, \lambda^2)$, $y(t) = a_1 y(t-1) + \dots + c_n e(t-n)$

(a_i's are assumed s.t. y SSP).

$\underbrace{\mathbb{E}[y(t)]}_{m_y} = a_1 \underbrace{\mathbb{E}[y(t-1)]}_{m_y} + \dots + c_n \underbrace{\mathbb{E}[e(t-n)]}_{\mu = m_e \neq 0}$

- Mean: (write the mean...)

$$m_y = \underset{\text{"DC gain"}}{W(1)} \cdot m_e = \frac{c_0 + c_1 + \dots + c_n}{1 - a_1 - \dots - a_m} \cdot \mu$$

• Cor fct: $\gamma_y(\tau) = \mathbb{E}\left[(y(t)-m_y)(y(t-\tau)-m_y)\right]$

$\boxed{\tau=0}$ $\mathbb{E}\left[(e(t)-\mu)^2\right] = \lambda^2+\mu^2$

$\boxed{\tau\neq0}$ $\mathbb{E}\left[e(t)e(t-\tau)\right] = \mu^2$.

computations very long.

TRICK:

$$\begin{cases} \tilde{y}(t) = y(t)-m_y = y(t)-W(1)\mu \\ \tilde{e}(t) = e(t)-\mu \end{cases}$$

Equation of $\tilde{y}(t)$ ←

$\equiv$

Equation of $y(t)$

$$\rightarrow \gamma_{\tilde{y}}(\tau) = \mathbb{E}\left[\tilde{y}(t)\tilde{y}(t-\tau)\right] = \mathbb{E}\left[(y(t)-m_y)(y(t-\tau)-m_y)\right] = \gamma_y(\tau).$$

$\tilde{y}$ has zero mean.

ANALYSIS IN THE FREQUENCY DOMAIN

How to analyze SSP in the frequency domain (not as easy as in deterministic world! Since you have $\neq$ realizations...)

SPECTRAL DENSITY OF SSP:

$$\Gamma_y(\omega) = \sum_{\tau=-\infty}^{\tau=+\infty} \gamma_y(\tau)\, e^{-j\omega\tau} \qquad (\omega \text{ frequency};\ j^2 = -1)$$

$$\shortparallel$$

$$F\{\gamma_y(\tau)\}$$: Discrete Fourier transform (DFT).

PROP OF $\Gamma_y(\omega)$:

- $\Gamma_y(\omega)$ is a REAL FCT of a REAL VARIABLE ω:

  $\forall \omega \in \mathbb{R},\ im(\Gamma_y(\omega)) = 0$.

- $\forall \omega \in \mathbb{R},\ \Gamma_y(\omega) \geq 0$ (positive).
- $\forall \omega \in \mathbb{R},\ \Gamma_y(-\omega) = \Gamma_y(\omega)$ (Even).
- $\forall \omega \in \mathbb{R}, \forall k \in \mathbb{Z},\ \Gamma_y(\omega) = \Gamma_y(\omega + 2k\pi)$ (2π-periodic).

⚠ $\omega \leq \omega_{max} = \pi$ (Nyquist freq.)

i.e $\frac{2\pi}{T} \leq \pi$ i.e $T \geq 2 = T_{min}$

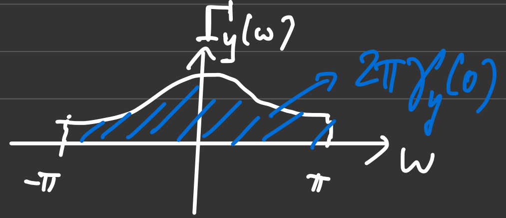
Inverse Transform:

$$\gamma_y(\tau) = F^{-1}(\Gamma_y(\omega)) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \Gamma_y(\omega)\, e^{j\omega\tau}\, d\omega$$

↳ $\gamma_y(\tau)$ & $\Gamma_y(\omega)$ carry the same info about y

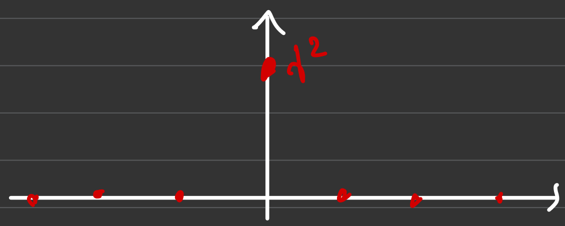
Rmk: $\gamma_y(0) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \Gamma_y(\omega)\, d\omega$

i.e

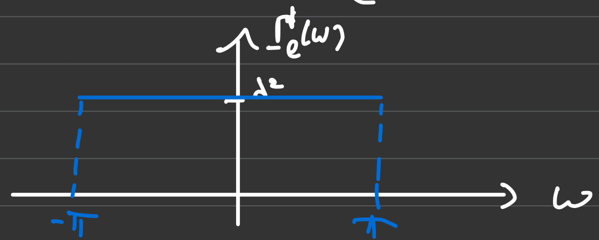


Example: Spectral density of $WN(\mu, \lambda^2)$:

$e(t) \sim WN(\mu, \lambda^2)$



$$\Gamma_e(\omega) = \sum_{\tau=-\infty}^{\tau=+\infty} \gamma_e(\tau)\, e^{-j\omega\tau} = \gamma_e(0) \cdot e^0 = \lambda^2 .$$



What about ARMA processes?

We cannot apply the definition of $\Gamma_y(\omega)$ because of the ∞ terms.

THM OF SPECTRAL FACTORIZATION:

$$v(t) \longrightarrow \boxed{W(z)} \longrightarrow y(t)$$

- $v(t)$ is SSP;
- $W(z)$ is asymptotically stable.

$$\Gamma_y(\omega) = \left| W(e^{j\omega}) \right|^2 \underbrace{\Gamma_v(\omega)}_{\text{if } v(t) \sim WN(\mu, \lambda^2),\ \Gamma_v(\omega) = \lambda^2 .}$$

N.B: $W(z)\big|_{z=e^{j\omega}} = W(e^{j\omega})$ is called "frequency response" of $W(z)$.

Different representations for ARMA processes:

1. Time domain: $y(t) = a_1 y(t-1) + \dots + c_0 e(t) + \dots + c_n e(t-n)$;
2. Operatorial: $y(t) = \frac{C(z)}{A(z)} e(t)$;
3. Probabilistic: $m_y, \gamma_y(\tau)$;
4. Frequency domain: $m_y, \Gamma_y(\omega)$.

equivalent representations.

Prediction

$\neq$ Representations of SSP **BUT**: Not unique.

How to define any SSP in a unique way?

↓

CANONICAL REPRESENTATION.

Theorem of Canonical Representation: let $y(t)$ be a SSP with $\Gamma_y(\omega)$ rational.

$\exists!$ pair $(W(z), e(t))$ s.t. $y(t) = W(z)e(t)$ if & only if, given $W(z) = \frac{C(z)}{A(z)}$

a) $C(z)$ and $A(z)$ are MONIC*;

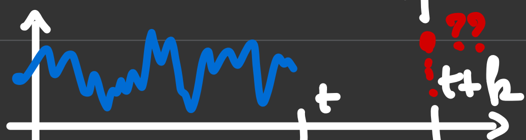
b) $C(z)$ and $A(z)$ have NULL relative degree;

c) $C(z)$ and $A(z)$ are COPRIME;

d) $C(z)$ and $A(z)$ have roots INSIDE unit circle.

* Leading Coeff = 1.

Question: given a SSP in canonical form & $\dots y(t-100), y(t-99), \dots y(t-1), y(t)$, how can we predict the future values of the process at time $t+k$?



WE WANT $\hat{y}(t+k|t)$ to be OPTIMAL in some sense (NEED an OPTIMALITY CRITERION) & we WANT to exploit ALL the information we have (model $\oplus$ data):

MODEL $\quad \xrightarrow{e(t)} \boxed{W(z)} \xrightarrow{y(t)}$

$\hat{y}(t+k|t)$ PREDICTOR

PAST DATA

$\dots y(t-100), \dots, y(t-1), y(t)$

<u>PREDICTION QUALITY:</u>

MEAN SQUARE PREDICTION ERROR: $\mathbb{E}\left[\left(y(t+k)-\hat{y}(t+k|t)\right)^2\right]$

$y(t+k)-\hat{y}(t+k|t) \triangleq \varepsilon(t+k|t) = \varepsilon_k(t)$

Question: Which predictor $\hat{y}(t+k|t)$ minimizes the criteria

$$\mathbb{E}\left[\varepsilon(t+k|t)^2\right] ?$$

$$\hat{y}(t+k|t) = f(\text{MODEL}, \text{PAST DATA})$$

HARD PB TO SOLVE IN THE GENERAL SETTING.

RESTRICTION: LINEAR PREDICTORS only.

Linear Predictors: $\hat{y}(t+k|t) = \sum_{i=0}^{+\infty} \alpha_i \, y(t-i) \quad \left(\text{s.t. } \sum_{i=0}^{+\infty} \alpha_i < +\infty \right).$

MODEL PAST DATA

$\rightarrow \hat{y}(t+k|t) = (\alpha_0 + \alpha_1 z^{-1} + \dots) y(t) = F_\alpha(z) \cdot y(t)$

$$y(t) \longrightarrow \boxed{F_\alpha(z)} \longrightarrow \hat{y}(t+k|t)$$

WE NEED TO FIND THE α_i's so that the mean squared prediction error is minimized: $\hat{\alpha} = \underset{\alpha_0, \alpha_1, \dots}{\operatorname{argmin}} \, \mathbb{E}\left[\left(y(t+k) - \hat{y}(t+k|t, \alpha) \right)^2 \right].$

SOLUTION OF DESIGN PROBLEM:

$$y(t) = W(z)e(t) \qquad e(t) \sim WN(0,\lambda^2)$$

$$y(t) = \sum_{i=0}^{+\infty} w_i\, e(t-i) \quad , \quad y(t-1) = \sum_{i=0}^{+\infty} w_i\, e(t-1-i) \ \ldots$$

$$\downarrow$$

$$\hat{y}(t+k|t) = \alpha_0\, y(t) + \alpha_1\, y(t-1) + \ldots$$

$$= \alpha_0 \sum_{i=0}^{+\infty} w_i\, e(t-i) + \alpha_1 \sum_{i=0}^{+\infty} w_i\, e(t-1-i) + \ldots$$

we can write it as

$$= \beta_0\, e(t) + \beta_1\, e(t-1) + \ldots = \sum_{i=0}^{+\infty} \underbrace{\beta_i}_{} e(t-i)$$

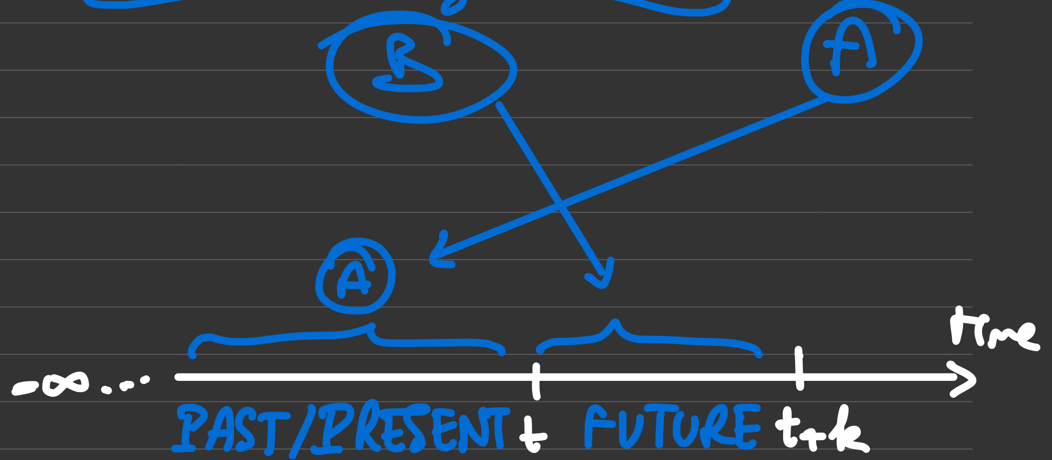
WE HAVE TO IDENTIFY β_i's.

REFORMULATION OF THE OPTI. PB IN TERMS OF β_i's :

$$\min_{\beta_0,\beta_1,\beta_2 \ldots} \mathbb{E}\left[\left(y(t+k) - \hat{y}(t+k|t)\right)^2\right]$$

: WAY EASIER TO SOLVE!

$$y(t+k) = \sum_{i=0}^{+\infty} w_i e(t+k+i) = \underbrace{\sum_{j=0}^{k-1} w_j e(t+k-j)}_{B} + \underbrace{\sum_{i=0}^{+\infty} w_{i+k} e(t-i)}_{A}$$



So:

$$\mathbb{E}\left[\left(y(t+k)-\hat{y}(t+k\mid t)\right)^2\right] = \mathbb{E}\left[\left(\sum_{j=0}^{k-1} w_j e(t+k-j) + \sum_{i=0}^{+\infty} w_{i+k} e(t-i) - \sum_{i=0}^{+\infty} \beta_i e(t-i)\right)^2\right]$$

$$= \mathbb{E}\left[\left(\text{NOT A FCT OF } \beta \,!\right)^2 + \left(\sum_{i=0}^{+\infty} (w_{i+k}-\beta_i) e(t-i)\right)^2 + 2\underbrace{\left(\cancel{\text{double cross product}}\right)}_{\text{all } e(t)\text{'s are uncorrelated}}\right]$$

OPTIMAL PREDICTOR GIVEN BY:

$$\left(\sum_{i=0}^{+\infty} (w_{k+i}-\beta_i) e(t-i)\right)^2 = 0$$

$\Updownarrow$

$\forall i = 0, 1, 2, \ldots \quad , \quad \beta_i = w_{k+i}$

OPTIMAL PREDICTOR OF THE NOISE:

$$\hat{y}(t+k|t) = \sum_{i=0}^{+\infty} w_{k+i}\, e(t-i)$$

ANALYSIS OF OPT. PREDICTORS

- LIMITATIONS OF THE PREVIOUS PREDICTOR:
  - depends on the past samples of $e(t)$;
  - described by an ∞ amount of w's;
  - we assumed $e(t)$ to be zero-mean.
- OPERATORIAL REPRESENTATION: $W(z) = \boxed{\dfrac{C(z)}{A(z)} = E[z] + \dfrac{z^{-k}F(z)}{A(z)}}$

$$y(t+k) = \frac{C(z)}{A(z)} e(t+k) = \left[E(z) + \frac{z^{-k}F(z)}{A(z)}\right] e(t+k)$$

$$= \underbrace{E(z)\, e(t+k)}_{} + \boxed{\frac{F(z)}{A(z)} \underbrace{z^{-k} e(t+k)}_{e(t)}}$$

Future samples of $e(t)$ = PREDICTION ERROR $\varepsilon(t+k|t)$

computable at time t: PREDICTOR $\hat{y}(t+k|t)$.

- $y(t+k)$ = PREDICTOR AT TIME t + PREDICTION ERROR.
- $\hat{y}(t+k|t) = \frac{F(z)}{A(z)} e(t)$: ARMA PROCESS.

PB: $\hat{y}$ depends on e :

WE can RECONSTRUCT e from $y(t), y(t-1), \dots$

$$y(t) = W(z) e(t) = \frac{C(z)}{A(z)} e(t)$$

$$\boxed{e(t) = \frac{A(z)}{C(z)} y(t)}$$

"WHITENING FILTER".

$$\hat{y}(t+k|t) = \frac{F(z)}{\cancel{A(z)}} \cdot \frac{\cancel{A(z)}}{C(z)} y(t)$$

$$= \frac{F(z)}{C(z)} y(t) : \text{FROM THE DATA !}$$

$\hat{y}(t+k|t, s)$ is a Stoch. Process ⊕ it is STATIONARY because $C(z)$ has roots inside the unit circle.

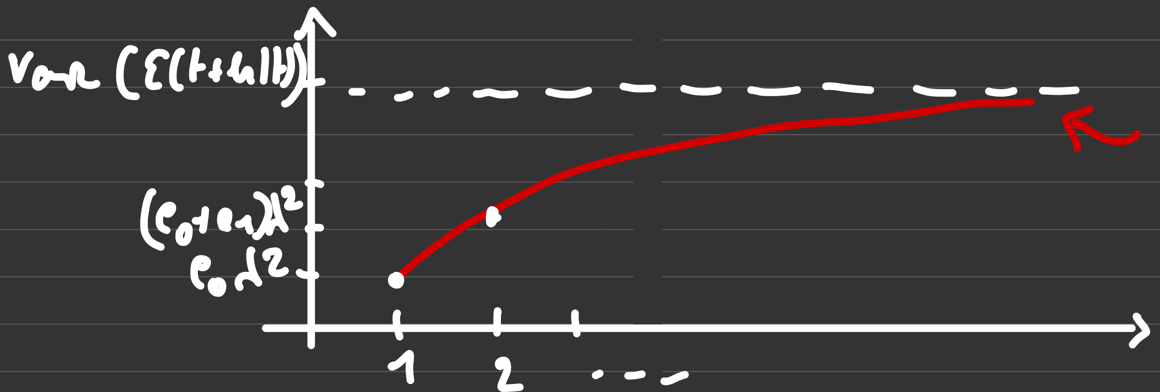
$$\boxed{\hat{y}(t+k|t) = \frac{F(z)}{C(z)}\, y(t)}$$

The quality of the pred gets worse when increasing k.

$$\varepsilon(t|t-k) = E(z)e(t) = e_0 e(t) + \dots + e_{k-1} e(t-k+1)$$

$$e(t) \longrightarrow \boxed{E(z)} \xrightarrow{\varepsilon(t|t-k)} \qquad (e(t) \sim WN(0,\lambda^2))$$

$$\mathbb{E}[\varepsilon(t+k|t)^2] = \mathbb{E}[\varepsilon(t|t-k)^2] = (e_0^2 + e_1^2 + \dots)\lambda^2$$



Asymptote where $\sum_{i=0}^{+\infty} e_i^2 \lambda^2 = Var(MA(\infty)) = Var(y(t))$.

- What if the initial process has $\neq 0$ mean?

$\hat{y}(t+k|t) = ?$

- REMOVE the mean from data: $\tilde{y}(t) = y(t) - m_y$
- COMPUTE $\hat{\tilde{y}}(t+k|t)$
- $\hat{y}(t+k|t) = m_y + \hat{\tilde{y}}(t+k|t)$ (ADD again m_y ...).

- What about ARMAX Processes?

$$y(t) = \underbrace{\frac{C(z)}{A(z)} e(t)}_{\text{stochastic}} + \underbrace{\frac{B(z)}{A(z)} u(t-d)}_{\text{Fully deterministic}} \quad : d = \text{pure delay.}$$

⇝ ASSUME : - $C(z)$ & $A(z)$ are canonical;
- $u(t)$ known from $t \to -\infty$ to $t \to +\infty$.

Define :

$$z(t) = y(t) - \frac{B(z)}{A(z)} u(t-d) = \frac{C(z)}{A(z)} e(t)$$

$$\hat{z}(t+k|t) = \frac{F(z)}{C(z)} z(t)$$

$$y(t) = \frac{B(z)}{A(z)} u(t-d) + z(t)$$

$$\hat{y}(t+k|t) = \frac{B(z)}{A(z)} u(t+k-d) + \hat{z}(t+k|t).$$

$$\overset{!}{=} \frac{B(z)}{A(z)} u(t+k-d) + \frac{F(z)}{C(z)} z(t)$$

$$= \frac{B(z)}{A(z)} u(t+k-d) + \frac{F(z)}{C(z)} \left(y(t) - \frac{B(z)}{A(z)} u(t-d) \right)$$

$$= \frac{B(z)}{C(z)} \underbrace{\left(\frac{C(z)}{A(z)} - \frac{F(z)}{A(z)} z^{-k} \right)}_{E(z)} u(t+k-d) + \frac{F(z)}{C(z)} y(t)$$

$$\boxed{\frac{C(z)}{A(z)} = E(z) + z^{-k} \frac{F(z)}{A(z)}}$$

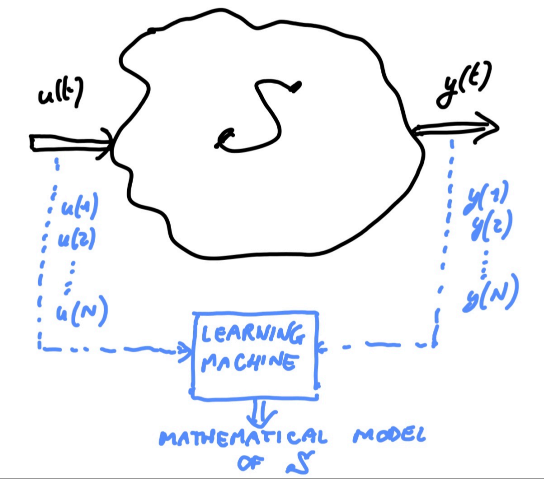
So:

$$\hat{y}(t+k|t) = \frac{B(z)E(z)}{C(z)} u(t+k-d) + \frac{F(z)}{C(z)} y(t)$$

<u>Depends on:</u>

- past data of $u(t)$ & $y(t)$ ✓
- model information ✓

MIDAI : IDENTIFICATION



PARAMETRIC SYSTEM IDENTIFICATION

1. Experiment design (and data collection)
2. Selection of a PARAMETRIC model structure $\mathcal{M}(\theta)=\{M_\theta : \theta\in\Theta\}$
3. Choice of the identification criterion $J_N(\theta)\geq 0$
4. Minimization of $J_N(\theta)$ wrt θ and computation of $\hat{\theta}_N = \arg\min_\theta J_N(\theta)$
5. Model validation

①

- Choice of $u(t)$: informativity of the dataset.
- Choice of experiment length N. confidence of the result.

②

- Choice of model class: $\mathcal{M}(\theta)=\{M_\theta : \theta\in\Theta\}$. Options → discrete time / C° time ; linear / non-linear ; time inv. / time varying ; static / dynamical

↳ ARMA/ARMAX MODELS: $y(t)=a_1y(t-1)+\dots+b_0u(t-d)+\dots+c_0e(t)+c_1e(t-1)+\dots$

↳ HERE: $\theta=[a_1\ a_2\ \cdots\ b_0\ \cdots\ c_0\ c_1\ \cdots]^T$: θ is $\boxed{\text{NOT}}$ sufficient, we need also

- $e(t)\sim WN(\mu,\lambda^2)$.
- pure delay d between IN/OUT PUT } $\boxed{\text{FOR FREE}}$.
- n_a, n_b, n_c : CRITICAL [we assume for now that they are given].

↳ Θ : set of admissible values for θ. E.g $y(t)=ay(t-1)+e(t)$ AR(1), $|a|<1$ } $\mathcal{M}(\theta)=\{AR_\theta(1), \theta=a, |a|<1\}$.

③

- Choice of $J_N(\theta)\geq 0$: "PREDICTIVE APPROACH"

$$J(\theta)=\mathbb{E}\left[\left(y(t+1)-\hat{y}(t+1|t,\theta)\right)^2\right]$$

IDEAL OBJECTIVE: a good model should return a low variance of the 1-step ahead prediction error

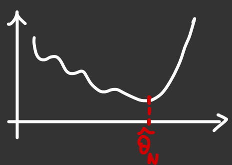
$$J_N(\theta)=\frac{1}{N}\sum_{t=1}^{N}\left(y(t)-\hat{y}(t|t-1,\theta)\right)^2$$

SAMPLE VERSION OF THE OBJECTIVE

why "1-step ahead predictor" of $y(t)$?
↳ if $\hat{y}(t+1|t)$ is the optimal pred. of $y(t)$, the variance of $\varepsilon(t+1|t)$ is: $Var[\varepsilon(t+1|t)]=Var[e(t)]=\lambda^2$. We obtain $\hat{\lambda}^2$ (an estimate of λ^2, better!) FOR FREE.

④

- Minimization of $J_N(\theta)$ wrt θ. $J_N(\theta) : \mathbb{R}^{n_\theta} \longrightarrow \mathbb{R}^+$. E.g $n_\theta = 1$



2 situations
- $J_N(\theta)$ quadratic : AR, ARX.
- $J_N(\theta)$ non-quadratic : ARMA, MA, ARMAX.

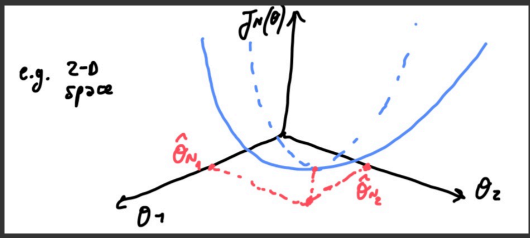
⑤

- Model validation. We made a few assumptions

We made a few assumptions
- S lies within a specific model set
- n_a, n_b, n_c fixed a-priori
- data could have been given by somebody else (could be non-informative)

WE NEED A QUALITY CHECK

IDENTIFICATION OF AR/ARX MODELS:



$$\mathcal{M}(\theta): \quad y(t) = \frac{B(z)}{A(z)} u(t-d) + \frac{1}{A(z)} e(t), \quad e(t) \sim WN(0,\lambda^2)$$

$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_m z^{-m}$$

$$B(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots b_{p-1} z^{-p+1}$$

⚠ $C(z) = 1$ (ARX model! No MA part)

$\theta = [a_1 \; a_2 \dots a_m \; b_0 \; b_1 \dots b_{p-1}]^T$ dimension $n_\theta = m+p$

$$\mathcal{D}_N = \{(u(1),y(1)), (u(2),y(2)), (u(3),y(3)) \dots (u(N),y(N))\}$$

$$\mathcal{M}(\theta): \quad A(z)\, y(t) = B(z)\, u(t-d) + e(t) \quad \Big) + y(t) - y(t)$$

$$y(t) - y(t) + A(z) y(t) = B(z) u(t-d) + e(t)$$

$$y(t) = (1 - A(z)) y(t) + B(z) u(t-d) + e(t)$$

$$y(t) = \underbrace{a_1 y(t-1) + a_2 y(t-2) + \dots a_m y(t-m) + b_0 u(t-d) + \dots b_{p-1} u(t-d-p+1)}_{\text{available at time } t-1} + \underbrace{e(t)}_{\text{not available at time } t-1}$$

$$J_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \left(y(t) - \hat{y}(t|t-1,\theta)\right)^2$$

$\hat{y}(t|t-1,\theta) = y(t) - e(t), \quad \varepsilon(t+1|t) = e(t)$

$$\tilde{\mathcal{M}}(\theta): \hat{y}(t|t-1,\theta) = a_1 y(t-1) + \dots + b_0 u(t-d) + \dots + b_{p-1} u(t-d-p+1)$$

Regression vector ["Regressor"]:

$$\varphi(t) = [y(t-1) \; y(t-2) \dots y(t-m) \; u(t-d) \dots u(t-d-p+1)]^T$$

$\Downarrow$

$$\boxed{\hat{y}(t|t-1,\theta) = \theta^T \varphi(t) = \varphi(t)^T \theta}$$

1 step ahead pred. in "vector form".

Linear function of θ.

$J_N(\theta)$ is a quadratic form of θ:

$$J_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \left(y(t) - \hat{y}(t|t-1,\theta)\right)^2$$

$$J_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \left[y(t) - \varphi(t)^T \theta\right]^2$$

(cf example in 2D ↗)

↳ Conditions to find $\hat\theta_N$:

1) $\hat\theta_N$ is a stationary point of $J_N(\theta)$: $\left.\frac{\partial J_N(\theta)}{\partial \theta}\right|_{\theta=\hat\theta_N} = 0$.

2) $\hat\theta_N$ is a minimum point: $\left.\frac{\partial^2 J_N(\theta)}{\partial \theta^2}\right|_{\theta=\hat\theta_N} > 0$.

$$\frac{\partial J_N(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial J_N(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial J_N(\theta)}{\partial \theta_{n_\theta}} \end{bmatrix} : \text{column vector.}$$

$$\frac{\partial J_N(\theta)}{\partial \theta} = \frac{d}{d\theta}\left[J_N(\theta)\right] = \frac{d}{d\theta}\left[\frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \varphi(t)^T\theta\right)^2\right] = \frac{1}{N}\sum_{t=1}^{N} \frac{d}{d\theta}\left[\left(y(t) - \varphi(t)^T\theta\right)^2\right]$$

$$\frac{\partial J_N(\theta)}{\partial \theta} = -\frac{2}{N}\sum_{t=1}^{N} \varphi(t)\left(y(t) - \varphi(t)^T\theta\right)$$

IDENTIFICATION OF ARMA/ARMAX MODELS:

$$\mathcal{M}(\theta): \quad y(t) = \frac{B(z)}{A(z)} u(t-d) + \frac{C(z)}{A(z)} e(t), \quad e(t) \sim WN(0, \lambda^2)$$

$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_m z^{-m}$$

$$B(z) = b_0 + b_1 z^{-1} + \dots b_{p-1} z^{-p+1}$$

⚠ $C(z) = 1 + c_1 z^{-1} + c_2 z^{-2} + \dots c_n z^{-n}$

$$\theta = [a_1 \; a_2 \; a_3 \dots a_m \; b_0 \; b_1 \dots b_{p-1} \; c_1 \; c_2 \dots c_n]^T$$

$$n_\theta = m + p + n$$

$$J_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \Big(y(t) - \underbrace{\hat{y}(t|t-1, \theta)}_{?} \Big)^2$$

$$\begin{array}{r|l} \overset{1 + c_1 z^{-1} + \dots}{C(z)} & \overset{1 - a_1 z^{-1} - \dots}{A(z)} \\ \hline - A(z) & 1 \\ \hline C(z) - A(z) & \end{array}$$

$$E(z) = 1$$

$$z^{-k} F(z) = C(z) - A(z)$$

BECAUSE WE WANT THE 1-STEP AHEAD PREDICTOR

↳ $k = 1$

$$\hat{y}(t|t-1) = \frac{F(z)}{C(z)} \cdot \underbrace{z^{-1} y(t)}_{y(t-1)} + \frac{B(z)}{C(z)} u(t-d) = \frac{C(z) - A(z)}{C(z)} y(t) + \frac{B(z)}{C(z)} u(t-d)$$

$$\varepsilon(t|t-1, \theta) = y(t) - \hat{y}(t|t-1, \theta) = y(t) - \frac{C(z) - A(z)}{C(z)} y(t) - \frac{B(z)}{C(z)} u(t-d)$$

$$= \left[1 - \frac{C(z) - A(z)}{C(z)} \right] y(t) - \frac{B(z)}{C(z)} u(t-d)$$

$$= \frac{A(z)}{\boxed{C(z)}} y(t) - \frac{B(z)}{\boxed{C(z)}} u(t-d)$$

nonlinear dependence on $c_1, c_2 \dots c_n$!

$$J_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t|t-1, \theta)^2$$

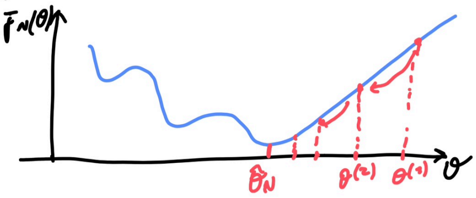
← no longer quadratic in θ ⚠

NONLINEAR OPTIMIZATION PROBLEM

We need an iterative (numerical) approach:

- We initialize the algorithm with an initial estimate $\theta^{(1)}$ (e.g. random choice)
- We define an UPDATE RULE $\theta^{(i+1)} = f(\theta^{(i)})$

The sequence of the estimates should converge to $\hat{\theta}_N = \arg\min_\theta J_N(\theta)$



2 main problems:
- many local optima
- update rule?

Problem of local minima

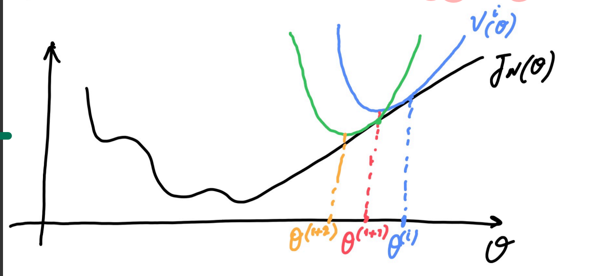
- Iterative algorithms are guaranteed to converge to a minimum, BUT it might be only a local one
- Several countermeasures (empirical approaches):

$$\left.\begin{array}{l} 1)\; \theta_1^{(1)} \to \theta_1^{(2)} \to \theta_1^{(3)} \dots \to \theta_1^{(N)} \\ 2)\; \theta_2^{(1)} \to \theta_2^{(2)} \to \theta_2^{(3)} \dots \to \theta_2^{(N)} \\ \vdots \\ M)\; \theta_M^{(1)} \to \theta_M^{(2)} \to \theta_M^{(3)} \dots \to \theta_M^{(N)} \end{array}\right\}$$

several solutions ... we take the one corresponding to minimum $J_N(\theta)$

The higher M, the higher the computational cost

Update rule for ARMAX models : Newton's method



$V^i(\theta)$ ... slide 8/22

[cf the computations]

MIDA1 : Model Validation :

$\mathcal{D}_N = \{(u(1),y(1)), \dots, (u(N),y(N))\}$
$\mathcal{M}_\theta = \{M(\theta) : \theta \in \Theta \in \mathbb{R}^{n_\theta}\}$
$\hat{\theta}_N = \arg\min_\theta J_N(\theta)$ with $J_N(\theta) = \frac{1}{N}\sum_{n=1}^{N} \overbrace{\varepsilon(t|t-1,\theta)^2}^{\varepsilon(t|t-1,\theta) = y(t) - \hat{y}(t|t-1,\theta)}$

Is $M(\hat{\theta}_N)$ a "good" model for $(y(t))_t$?

For $N \to +\infty$, all $J_N(\theta)$ CV to a SINGLE COST.

Theorem

Under the current assumptions, as $N \to +\infty$

$$J_N(\theta, s) \xrightarrow[N \to +\infty]{} \bar{J}(\theta) = \mathbb{E}_s[\varepsilon(t|t-1,\theta,s)^2]$$

Moreover, by letting

$$\Delta = \{\theta^*, \bar{J}(\theta^*) \leq \bar{J}(\theta), \forall \theta\}$$

be the set of global minimum points of $\bar{J}(\theta)$, we have that

$$\hat{\theta}_N(s) \xrightarrow[N \to +\infty]{} \Delta \quad \text{w.p. } 1$$

Corollary: If $\Delta = \{\theta^*\}$, i.e. $\bar{J}(\theta)$ has a unique minimum, then $\hat{\theta}_N(s) \xrightarrow[N \to +\infty]{} \theta^*$ w.p. 1.

Model Order Selection :

Let $\mathcal{S}$ be the R system. 4 cases
- $\mathcal{S} \in \mathcal{M}(\theta)$ or not.
- Δ is a singleton or not.

$\mathcal{M}(\theta) = \{M(\theta) : \theta \in \Theta \subseteq \mathbb{R}^{n_\theta}\}$ Fixed-order model class so far.

We cannot guarantee that $\mathcal{S} \in \mathcal{M}(\theta)$ & we don't know how to enlarge it if $\mathcal{S} \notin \mathcal{M}(\theta)$.

For simplicity, we'll consider $n = m = p$ (single scalar parameter). How to select n ?

- If we just increase $n \nearrow n_{max}$, then $J_N^{(n)}(\theta)$ is $\searrow$ w/ n but hard to distinguish under/over fitting.

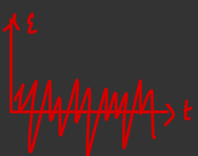
$\rightarrow$ 3 criteria for model selection :
1) Whiteness test on residuals.
2) Cross-validation.
3) Identification with model order penalties.

1) Whiteness test on residuals :

"Anderson's test"

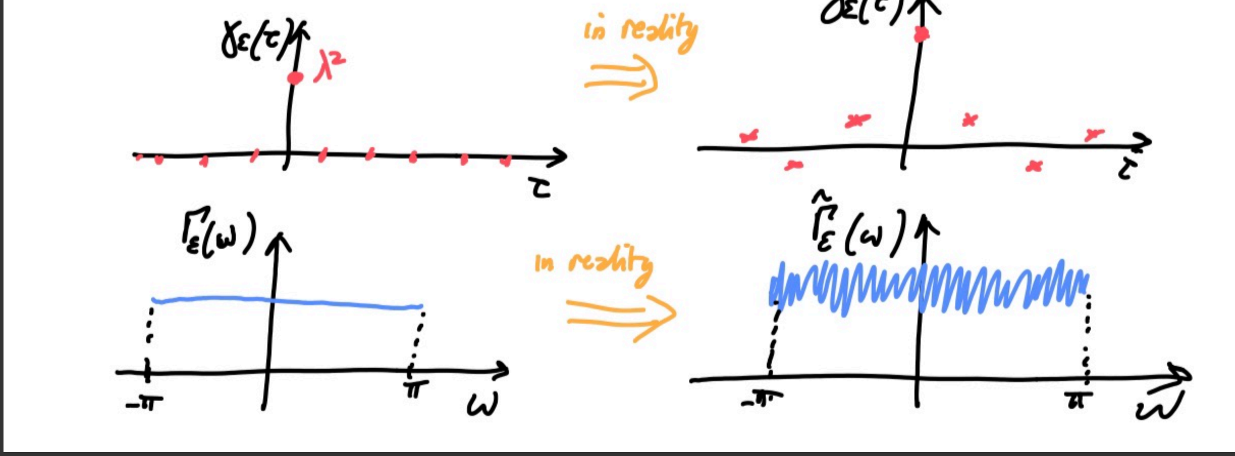
We know that if $\mathcal{S} \in \mathcal{M}(\theta)$, then, for a certain n we have $\underbrace{\hat{\theta}_N^{(n)} \simeq \theta^0}_{\mathcal{S} = M(\theta^0)}$ and $\varepsilon(t|t-1,\hat{\theta}_N^{(n)}) \simeq \underbrace{e(t)}_{WN(0,\lambda^2)}$.

So the question is: given $\varepsilon(t|t-1,\hat{\theta}_N^{(n)}) = y(t) - \hat{y}(t|t-1,\hat{\theta}_N^{(n)})$

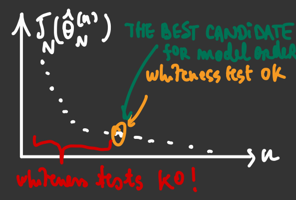


: Is it the realization of a WN ?

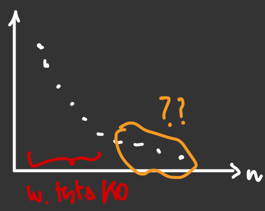
Whiteness test on $\varepsilon(t|t-1,\theta)$



Ideal Situation:



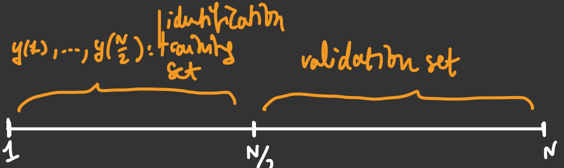
Real Situation:



2) Cross-validation:

$$D_N = \{(u(1),y(1)),\dots,(u(N),y(N))\}$$

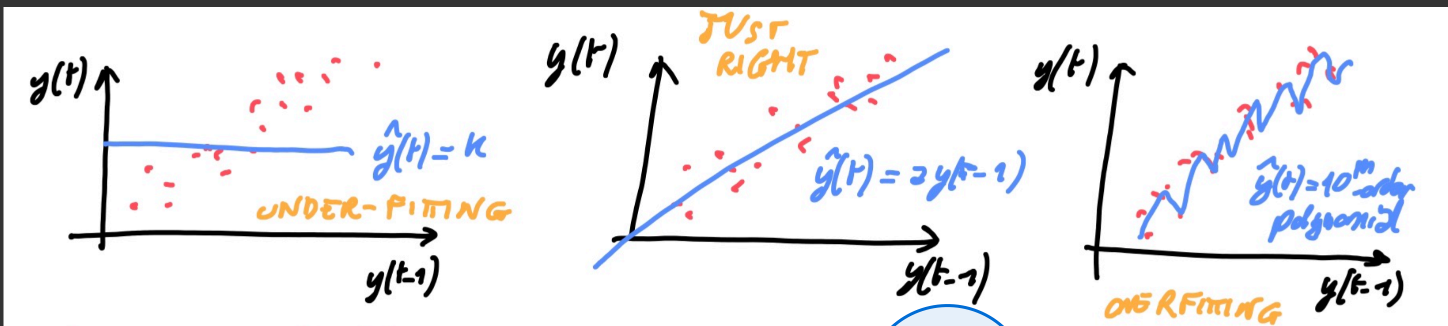
$$J_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t)-\hat{y}(t|t-1,\theta)\right)^2$$



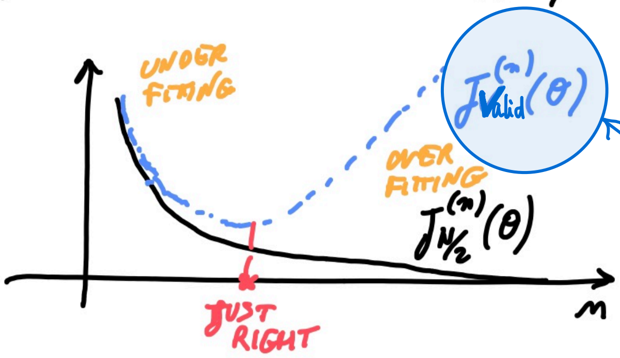
Procedure:

- for $n=1$ to $n=n_{max}$, find $\hat{\theta}_{\frac{N}{2}}^{(n)} = \underset{\theta}{\operatorname{argmin}}\; J_{N/2}^{(n)}(\theta) = \underset{\theta}{\operatorname{argmin}}\; \frac{1}{(N/2)}\sum_{t=1}^{N/2}\left(y(t)-\hat{y}(t|t-1,\theta)\right)^2$

  And evaluate $J_{valid}\left(\hat{\theta}_{N/2}^{(n)}\right) = \frac{N}{2}\sum_{n=\frac{N}{2}+1}^{N}\left(y(t)-\hat{y}(t|t-1,\hat{\theta}_{N/2}^{(n)})\right)^2$.
- choose n (optimal model order) as the one which minimizes $J_{valid}\left(\hat{\theta}_{N/2}^{(n)}\right)$.



S: $y(t) = a\,y(t-1) + e(t)$



Not the same shape as the cost on the training set!

3) Identification costs with order penalties:

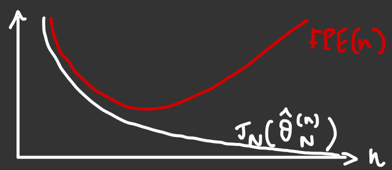
Cross validation : same cost for training / validation, ≠ data.
Penalized - cost approach : ≠ cost for model order, same data.

} CV ≠ Penalization.

↳ Instead of minimizing $J_N^{(n)}(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t)-\hat{y}^{(n)}(t/t-1,\theta)\right)^2$, we minimize:

① Final Prediction Error.
② Akaike I.C (AIC).
③ Minimum Desc. length.

① Final Prediction Error : (FPE)



$\min_{\theta} E\left[J_N(\theta)\right]$ ideal objective (UNFEASIBLE).

$FPE(n) = \frac{N+n}{N-n} J_N(\hat{\theta}_N^{(n)})$ "feasible proxy" of the ideal objective.

② AIC :

$$AIC(n) = \underbrace{\ln\left(J_N(\hat{\theta}_N^{(n)})\right)}_{\searrow \text{ with } n.} + \underbrace{2\,\frac{n}{N}}_{\text{Penalty on } n.}$$

FPE vs AIC ? $\ln(FPE) \simeq \frac{n}{N} - \left(-\frac{n}{N}\right) + \ln\left(J_N(\hat{\theta}_N^{(n)})\right) \simeq AIC$

→ FPE & AIC return the same optimal model order.

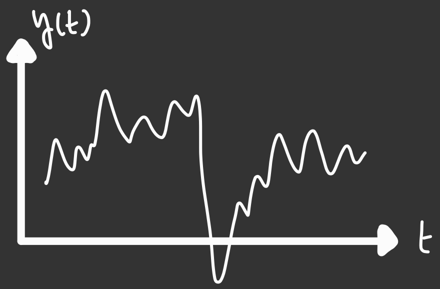
③ MDL :

$$MDL(n) = \underbrace{\ln\left(J_N(\hat{\theta}_N^{(n)})\right)}_{\searrow \text{ with } n.} + \underbrace{\ln(N)\frac{n}{N}}_{\text{Penalty on } n.}$$

AIC vs MDL ? $\ln(N) > 2$ for $N > 8$, so the model order is + penalized by MDL.

In the general case, $\{ S \notin \mathcal{M}(\theta) \}$ we usually prefer to slightly overfit & then we use AIC.

MIDA I : Non Parametric Estimation :



So far, when we wanted to infer smthg about the SI, we had to learn the model!

$m_y = \mathbb{E}[y(t)] \rightsquigarrow$ learn $W(z) = W(z,\theta)$ and then : $\hat{m}_y = W(1, \hat{\theta}_N) \cdot m_e \dots$

$e(t) \longrightarrow \boxed{W(z)} \longrightarrow y(t)$

It is possible to estimate m_y, $\gamma_y(\tau)$, $\Gamma_y(\omega)$ directly from the DATA, without first identifying a full model of $W(z)$!

DEF (CORRECTNESS): An estimator is correct if the expected value of the estimator is equal to the "probabilistic property to be estimated" (see example).

DEF (CONSISTENCY): An estimator is consistent if the estimate-error-variance tends to zero when the number of data goes to $+\infty$.

Thm about consistency: If $\gamma_y(\tau) \xrightarrow[|\tau| \to +\infty]{} 0$ then $\left|\begin{matrix} \hat{m}_N \\ \hat{\gamma}_N(\tau) \end{matrix}\right.$ is consistent.

- **SAMPLE MEAN:** $D_N = \{y(1), y(2), \dots, y(N)\} \xrightarrow[\text{PATH}]{\text{DIRECT}} \mathbb{E}[y(t)]$

NATURAL ESTIMATOR : $\hat{m}_N = \frac{1}{N} \sum_{t=1}^{N} y(t)$

CORRECTNESS OF $\hat{m}_N$: YES $\longrightarrow \mathbb{E}[\hat{m}_N] = \mathbb{E}\left[\frac{1}{N}\sum_{t=1}^{N} y(t)\right] = \frac{1}{N}\sum_{t=1}^{N} \mathbb{E}[y(t)] = \frac{1}{\cancel{N}} \cancel{N}\, \mathbb{E}[y(t)]$ □

CONSISTENCY OF $\hat{m}_N$: NO $\longrightarrow \mathbb{E}\left[(\hat{m}_N - m_y)^2\right] \underset{\uparrow}{=} \mathbb{E}\left[\left(\frac{1}{N}\sum_{t=1}^{N} y(t,s) - m_y\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{t=1}^{N} v(s) - 0\right)^2\right] = \mathbb{E}\left[v(s)^2\right] = 1$ □

e.g. $y(t,s) = v(s)$ where $v(s) \sim \mathcal{N}(0,1)$.

But yes for all ARMA processes (cf thm above).

- **SAMPLE COVARIANCE FUNCTION :** Without loss of generality, we assume $m_y = 0$.

NATURAL ESTIMATOR : $\hat{\gamma}_N(\tau) = \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} y(t)\, y(t+|\tau|)$, $0 \leq \tau \leq N-1$.
(Practically : $\tau << N-1$)

CORRECTNESS OF $\hat{\gamma}_N$: YES $\longrightarrow \mathbb{E}\left[\frac{1}{N-\tau}\sum_{t=1}^{N-\tau} y(t) y(t+\tau)\right] = \frac{1}{N-\tau}\sum_{t=1}^{N-\tau} \gamma_y(\tau) = \gamma_y(\tau)$. □

CONSISTENCY OF $\hat{\gamma}_N$: YES For all ARMA processes (cf. Thm above).

• **SAMPLE SPECTRUM:** $\Gamma_y(\omega) = \sum_{\tau=-\infty}^{\tau=+\infty} \gamma(\tau) e^{-j\omega\tau}$

POSSIBLE ESTIMATOR: $\hat{\Gamma}_N(\omega) = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}_N(\tau) e^{-j\omega\tau}$ 1st source of approximation.

2nd source of approximation.

CORRECTNESS OF $\hat{\Gamma}_N(\omega)$: ONLY ASYMPTOTICALLY CORRECT $\rightarrow$ $\mathbb{E}[\hat{\Gamma}_N(\omega)] = \sum_{\tau=-(N-1)}^{N-1} \mathbb{E}[\hat{\gamma}_N(\tau)] e^{-j\omega\tau}$

$$= \sum_{\tau=-(N-1)}^{N-1} \gamma_y(\tau) e^{-j\omega\tau} \xrightarrow[N\to+\infty]{} \Gamma_y(\omega). \square$$

CONSISTENCY OF $\hat{\Gamma}_N(\omega)$: NO $\rightarrow$ $\mathbb{E}\left[\left(\hat{\Gamma}_N(\omega) - \Gamma_y(\omega)\right)^2\right] \longrightarrow \Gamma(\omega)^2 \neq 0$ in general. $\square$

Remark: $\mathbb{E}\left[\left(\hat{\Gamma}_N(\omega_1) - \Gamma_y(\omega_1)\right)\left(\hat{\Gamma}_N(\omega_2) - \Gamma_y(\omega_2)\right)\right] \xrightarrow[N\to+\infty]{} 0$

FINAL OBSERVATION: we can compute the spectrum like:

$\hat{\Gamma}'_N(\omega) = \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{-j\omega t} \right|^2$ * which is proven to be equivalent to:

$\hat{\Gamma}_N(\omega) = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}'_N(\tau) e^{-j\omega\tau}$ with: $\hat{\gamma}'_N(\tau) = \frac{1}{\boxed{N}} \sum_{t=1}^{N-\tau} y(t) y(t+\tau)$.

* Easy to compute (DFT...).

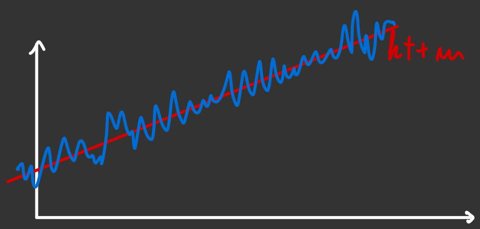
MIDA I: DATA PRE PROCESSING:

If $y(t)$ non-stationary? $\rightarrow$ trend / seasonality

we 1st need to estimate the non-stationarity, to remove it & work with the remainder SSP.

• TREND REMOVAL:

$$y(t) = \underbrace{\tilde{y}(t)}_{SSP} + \underbrace{kt + m}_{\text{linear trend}}, \quad \forall t.$$



The only available information is:

$$D_N = \{y(1), \dots, y(N)\}$$

What if we want to compute $\hat{y}(t+1|t)$? We cannot apply the predictor formula for SSP to this data, since it is not stationary!

ALGORITHM: ① Estimate m & k characterizing the trend $\rightarrow (\hat{m}, \hat{k})$.

?

② $\tilde{y}(t) \cong y(t) - \hat{k}t - \hat{m}$

③ $\hat{\tilde{y}}(t+1|t)$ from the formula

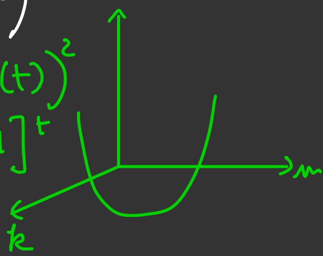
$\hat{y}(t+1|t) = \hat{\tilde{y}}(t+1|t) + \hat{k}(t+1) + \hat{m}$

ESTIMATE m & k : $\mathbb{E}\left[y(t) - kt - m\right] = \mathbb{E}\left[\tilde{y}(t)\right] = 0$

$$\downarrow$$

$$(\hat{m}, \hat{k}) = \underset{m,k}{\arg\min} \ \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - kt - m\right)^2$$

see the least squares results of ARX identification. } where $\theta = [k \quad m]^T$ & $\varphi(t) = [t \quad 1]^T$ $\left(y(t) - \theta^T\varphi(t)\right)^2$

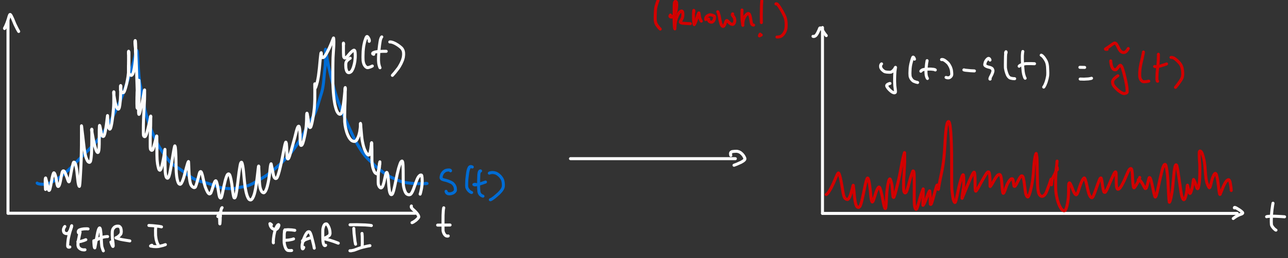


$$\rightarrow \begin{bmatrix} \hat{k} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{N} t^2 & \sum_{t=1}^{N} t \\ \sum_{t=1}^{N} t & N \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{N} t\, y(t) \\ \sum_{t=1}^{N} y(t) \end{bmatrix}$$

• SEASONALITY REMOVAL:

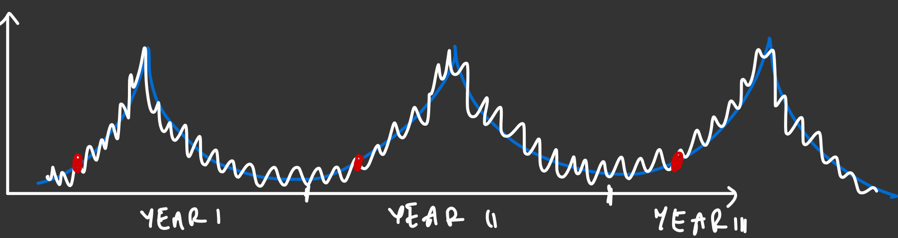
$y(t) = \underset{\text{SSP}}{\tilde{y}(t)} + \underset{\text{seasonality}}{s(t)}$, $\forall t$, with

assumption: $s(t) = s(t + kT)$, $k \in \mathbb{Z}$, T (period) (known!)



We first need to estimate $s(t)$, then compute $\hat{\tilde{y}}(t+1|t)$ and finally
$\hat{y}(t+1|t) = \hat{\tilde{y}}(t+1|t) + \hat{s}(t+1)$.

ESTIMATE $s(t)$:



$M \cdot T \leq N$

$$\begin{cases} \hat{s}(t) = \dfrac{1}{M}\sum_{k=0}^{M-1} y(t+kT), \quad t = 1, 2, \dots, T. \\ M: \text{nb. of available periods in the dataset.} \end{cases}$$

UNDERLYING IDEA : $\hat{s}(t) = \frac{1}{M}\sum_{k=0}^{M-1} y(t+kT) = \underbrace{\frac{1}{M}\sum_{k=0}^{M-1}\tilde{y}(t+kT)}_{} + \underbrace{\frac{1}{M}\sum_{k=0}^{M-1} s(t+kT)}_{}$

If M large, since the sample mean is consistent: $\xrightarrow[M\to+\infty]{} \mathbb{E}[\tilde{y}(t)] = 0$

$\forall k \in \mathbb{Z}$, $s(t) = s(t+kT)$ so: $= \frac{1}{\cancel{M}} \cdot \cancel{M} s(t) = s(t)$

so $\hat{s}(t) \longrightarrow s(t)$ if M large.

N.B: If T unknown, we can estimate it from the data.