27/9/18

Time Series Analysis

→ Regression ⇒ $y_i = f(\underset{\sim}{x}_i) + \epsilon_i$ ⟶ get a model
→ estimate model parameters
→ predict values of Y for given $\underset{\sim}{X}$.

$(y_i, \underset{\sim}{x}_i)$ ⇒ random sample of population BUT no such association with time.

→ Time series Analysis:

Indx set = T (time stamp set) corresponds to time.

Def^n: Time series is a collection of random variables over index set T (in the order of occurrence) which might be finite, countably infinite or uncountable set

⇒ $\boxed{\{X_t\} \text{ where } t \in T}$ → if countable ⇔ discrete TS.
→ if intervals ⇔ continuous TS

not properties/adjectives of $\{X_t\}$, disc/cont properties of T (time) only
real or adjectives

values of data taken a time point, t=i, i=1,2,3... ⟵ $X_1, X_2, X_3 \ldots$ may have diff. features.

Def^n: Time Series is a coll^n of RV $\{X_t \mid t \in T\}$ .. used for theoretical aspect.

Realized values: (i.e. dataset) else, these 2 notations are used interchangably.
$\{X_1 = x_1, X_2 = x_2 \cdots, X_n = x_n\}$, where x_i's are some numeric or categorial values.

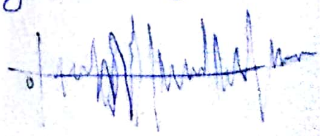
- White Noise: time series generated from uncorrelated variables with zero mean & fixed variance is called W.N.

$$W_t \sim WN(0, \sigma_w^2)$$

ex. $W_t = X_t \overset{i.i.d.}{\sim} WN = N(0,\sigma^2)$ (trivial example for strongly stationary)

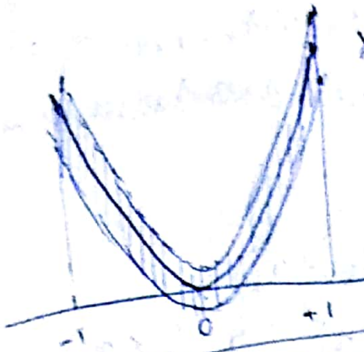
ex: $X_t = W_t = \begin{cases} N(0,1), & \text{if } t \text{ is even} \\ (Exp(1) - 1) & \text{if } t \text{ is odd} \end{cases}$

$\downarrow \lambda = 1$
$\lambda e^{-\lambda x} \Rightarrow e^{-x}$



$Y|X=x \sim U(x^2-0.5, x^2+0.5)$ ← but dependent.

Dependency & Correlation?



$X \sim U(-1,1) \Rightarrow E(X)=0, Var(X)=0$

$E(X,Y)=0$

$Cov(X,Y)=0$

$\therefore$ no corr.

- uncorr
- fixed mean

dependency doesn't come into picture → can be dependent

Seasonality (Repetitive part): Deterministic pattern which returns after a fixed interval.

$$\boxed{X_t = T_t + W_t + S_t}$$

analogous: $\underset{\sim}{y} = \underbrace{X\underset{\sim}{\beta}}_{\text{determ. part}} + \underbrace{\underset{\sim}{\varepsilon}}_{\text{R.E.}}$

ex $X_t = T_t + S_t + W_t = 0.3t + 3\sin(\frac{\pi t}{?}) + W_t$ or $= \sin(\frac{\pi t}{2}) + W_t$

- Additive Model
- Multiplicative Model. $X_t = T_t \times W_t \times S_t$

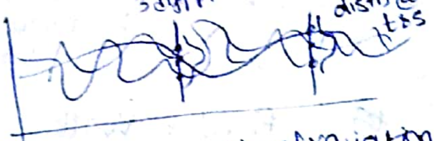
Trend: Deterministic pattern persist throughout entire T.S.

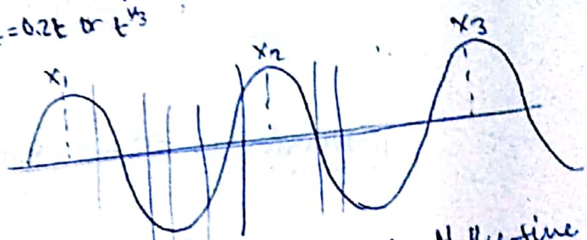
ex. $X_t = T_t + W_t$ where, $T_t = 0.2t$ or $t^{1/3}$

White Noise → Random error added to the deterministic part. (+ seasonality)?

$$\boxed{X_t = T_t + S_t + W_t}$$

3 diff pts at t's — distn's @ t's





$\Rightarrow X_1 = 1$ all the time

$\therefore$ frequency of X_i is important while doing TSA.

"Waiting" times

* There is not notion of variation for the entire time series (it can run from $-\infty$ to $+\infty$) $\therefore$ Point to point variance or variance at point t, $Var(X_t)$.

- Mean: Suppose $\{X_t\}$ is a T.S. with $E|X_t| < \infty$

  mean $f^n \Rightarrow \mu_t = E[X_t]$

- ACVF: Suppose that $\{X_t\}$ is a time series with $E[X_t^2] < \infty$ then its autocovariance function is

  comparing $t=t, t=s$ points in same TS.

  $$\gamma_X(s,t) = Cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)]$$

  means of the same T.S.

  multiple times, readings (X_1, X_2) ~~take at~~ ~~between~~

  Stationarity ⇒ constant mean & variance over the T.S. ⇒ no trend in series

- Weakly Stationarity (of T.S.) ⇒ constant auto-cov. as well.

  strongly Stationarity (of T.S.)

Strong Stationarity ⇒ Weak Stationarity.

"Random Walk"

If $F(Y_t) = F(Y_{t+h})$ ⇒ strict stationarity

28/9/18

Strongly Stationary T.S $\Rightarrow$ $\{x_t\}$ is strongly stationary if for all $k, h, t_1, \dots t_k, x_1, \dots x_k$ · shifting the time axis doesn't affect the ...

i.e.

$P(X_{t_1} \le x_1, \dots X_{t_k} \le x_k)$

$= P(X_{t_1+h} \le x_1, \dots X_{t_k+h} \le x_k)$

Weakly Stationary T.S, $\{X_t\}$ iff

1. μ_t is independent of t
2. for each $h \in \mathbb{Z}$, ACVF $\gamma(t+h, t)$ is independent of t.

ex. $X_t \sim WN(0, \sigma^2_w)$

NOT ex. $X_t = \sum_{i=0}^{t} W_i$, $W \sim WN(0,\sigma^2)$

ex. $X_t \sim WN = N(0, \sigma^2)$ ✓

Not an ex: Random Walk is defined as:

$S_t = \sum_{i=0}^{t} X_i$, where $X_i \overset{iid}{\sim} N(0,\sigma^2)$

i.e. x_i iid $N(0,\sigma^2)$

after 2D, there is no prob. that we can return starting point

$\boxed{\text{Strongly St. T.S} \Rightarrow \text{Weakly St. T.S.}} \Rightarrow$ iff, moment exists i.e. mean & var. d.n.e.

- $X_i \overset{iid}{\sim} N(0,\sigma^2)$ $i = 1, 2, \dots$

$X_0 = 0$

$S_t = \sum_{i=0}^{t} X_i$

$E(S_t) = ?$

$V(S_t) = ?$

autoCov$(S_t, S_s) = ?$

(let $t \le s$)

Random Walk:

$E(S_t) = E\left(\sum_{i=0}^{t} X_i\right) \Rightarrow \sum_{i=0}^{t} E(X_i) = \sum 0 = \underline{0}$

$V(S_t) = E\left(\sum X_i - E(\sum X_i)\right)^2 = E(S_t - E(S_t))^2$

$= E(S_t^2) - (E(S_t))^2$

$= E\left(\left(\sum_{i=0}^{t} X_i\right)^2\right) = E\left((X_1 + X_2 + \dots X_t)^2\right)$

$= (\sum E(X_i)^2) + \sum_{i \ne j} 2 \cdot E(X_i) E(X_j)$

$\because$ They are uncorrelated

$= 0 + 0 = \underline{0}$

autoCov$(S_t, S_s) = ?$

$\Rightarrow Cov(S_t, S_{t+0})$

$\Rightarrow Cov\left(\sum_{i=0}^{t} X_i, \sum_{i=0}^{t} X_i + \sum_{i=t}^{s} X_i\right)$

$\Rightarrow Cov\left(\sum_{i=0}^{t} X_i, \sum_{i=0}^{t} X_i\right) + \cancel{Cov\left(\sum_{i=0}^{t} X_i, \sum_{i=t}^{s} X_i\right)}^{0}$

$\Rightarrow \underline{\sigma^2 \cdot t \cdot 1}$

$\because$ uncorrelated

$\because$ def^n of ...

$Z_t = \mu t + S_t$ where, S_t is a random walk then Z_t is called random walk with drift (μ).

<u>Auto Regression:</u> $\Rightarrow$ "Regression among the same time series"

W_t is WN $|\phi|<1 \longrightarrow$ weakly st.

$X_t = \phi \cancel{X}_{t-1} + W_t$ = "Random Walk"

* NOT independent $W_i \overset{iid}{\sim} N(0,\sigma^2)$.

[$\because$ These terms are LC of eachother]

let $x_0 = 0$, $t \geq 1$

$X_1 = W_1$

$X_2 = \phi W_1 + W_2 = W_2 + \phi W_1$

$X_3 = W_3 + \phi W_2 + \phi^2 W_1$

$\vdots$

$$X_t = W_t + \sum_{i=0}^{t-1} \phi^{t-i} W_i = \underline{\underline{\sum_{i=0}^{t} \phi^{t-i} W_i}}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \sigma^2 \begin{bmatrix} 1 & \phi & \phi^2 & \cdots \\ \phi & 1 & \cdots & \\ \vdots & & \ddots & \\ \phi & \phi & \cdots & 1 \end{bmatrix}\right]$$

$E(X_t) = 0$

$V(X_t) = ?$

$$= E(X_t^2) - \overset{\nearrow 0}{(E(X_t))^2}$$

$$= E\left(\left(W_t + \sum_{i=0}^{t} \phi^{t-i} W_i\right)^2\right)$$

$$= (E(W_t))^2 + \left(E\left(\sum_{i=0}^{t} \phi^{t-i} W_i\right)\right)^2 + 2E(W_t)\cdot E\left(\sum_{i=0}^{t} \phi^{t-i} X_i\right)$$

$$= 0 + \sum \phi^{2(t-i)} E(X_i^2) + 2\underbrace{\sum_i \sum_j \phi^{t-i} \phi^{t-j} E(X_i)E(X_j)}_{\to 0} + 0 \quad [E(X_i^2) = \sigma^2]$$

$$V(X_t) = \sigma^2 \sum_{i=1}^{t-1} (\phi^{t-i})^2, \quad t \uparrow \infty$$

$$= \sigma^2 (1 + \phi^2 + \phi^4 + \cdots) = \frac{\sigma^2}{1-\phi^2}$$

* After some time there is no impact of starting point.

~ Markov...

~ Monte Carlo Sim.

$\Downarrow$

we apply multiple

~ RV, X_t are identically distributed.

~ $(X_t, X_{t+h}) \overset{d}{=} (X_1, X_{1+h})$ for all integers t & h

~ X_t is weakly stationary if, $E(X_t^2) < \infty \ \forall t$

~ Weak Stationarity does not imply strict stationarity.

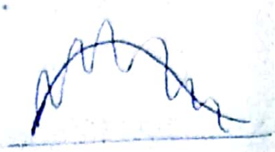
~ An i.i.d. sequence is strictly stationary.

* Even though all moments are same $\not\Rightarrow RV_1 = RV_2$ (not same)

* ① $f(x)$ pdf of lognormal(0,1)

② $(1+\sin(2\pi x))\, f(x)$

check if the distb's are same.



→ Operators:

① Back Shift Operator:

$$\boxed{B^h X_t = X_{t-h}}$$

ex: $BX_t = X_{t-1}$

$B^2 X_t = BX_{t-1} = X_{t-2}$

$\nabla \quad \nabla^2 \quad \dots \quad \nabla^n$

$P_n(x)$: $x_0 \to y_0$, $x_1 \to y_1$, $\vdots$, $x_k \to y_k$

$y_1 - y_0$, $y_2 - y_1$, $\vdots$

$y_2 - 2y_1 + y_0$

Const

error cannot be removed.

NO structure ONLY error @ ∇^n.

∴ ② Difference Operator:

$\nabla X_t = X_t - X_{t-1} = (I-B)X_t$ → "seasonal difference"

$I \Rightarrow$ identity

$$\boxed{\nabla = I - B}$$

$\Rightarrow \nabla^h X_t = (I-B)^h X_t$.

∴ Seasonal Difference, $\nabla_s X_t = (1-B^s)X_t$

ex: $\nabla = I - B$

$\nabla^2 = I - 2B + B^2$

$\nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}$

check if the distb'ns are same.

ift Operator:

$\boxed{\ldots_t = X_{t-h}}$

∇^n

const

error cannot be removed.
NO structure ONLY error @ ∇^n.

$\rightarrow$ "seasonal difference"

$I \Rightarrow$ identity

$)X_t$

→ Estimation of Trend in the Absence of Seasonality

1/10/18

• Smoothing with a finite moving avg. filter: (MA Method):

$$\begin{array}{lcl}
x_1 \\
x_2 \Big\} q & & NA \\
 & & NA \\
x_3 & \longrightarrow & \frac{1}{5}\sum_{i=-2}^{2} X_{3+i} \\
x_4 & \longrightarrow & \frac{1}{5}\sum_{i=-2}^{2} X_{4+i} \\
x_5 & & \\
\vdots & & \vdots \\
x_{48} & \longrightarrow & \frac{1}{5}\sum_{i=-2}^{2} X_{48+i} \\
x_{49} \Big\} q & & NA \\
x_{50} & & NA
\end{array}$$

$t=3$

sliding moving window

assume, $X_t = T_t + \cancel{S_t} + \cancel{W_t}$

$X_t = T_t + W_t$.

$$m_t = \frac{1}{2q+1}\left(\sum_{i=t-q}^{t+q} X_i\right) = \frac{1}{2q+1}\sum_{i=t-q}^{t+q} T_i \approx T_t$$

if T_t is linear in $[t-q, t+q]$

window length $= 2q+1$ (say) $= 5$

"Moving Average Method"
- can't do for first 'q' & last 'q' observations.

Linear Filter:

$$\boxed{m_t = \sum_{i=-q}^{q} a_i X_{t-i}}$$

multiple linear regression? or weighted avg.

Where, 'q' is a non-negative integer
two-side moving avg. is given by $\underline{\underline{m_t}}$

• Exponential Smoothing:

$*$ $m_t = X_1$ if $t=1$

$\Rightarrow m_1 = X_1$

$*$ $m_t = \alpha X_t + (1-\alpha) m_{t-1}$, $t \geq 2$

$\therefore m_2 = \alpha X_2 + (1-\alpha) m_1$

$= \alpha X_2 + (1-\alpha) X_1$

$m_3 = \alpha X_3 + (1-\alpha) m_2$

$= \alpha X_3 + (1-\alpha)\alpha X_2 + (1-\alpha)^2 X_1$

$\longleftarrow$ more wt. towards the present

$\longrightarrow$ less weight as we go to the past

$$\boxed{\therefore m_t = \alpha X_t + (1-\alpha) m_{t-1} \quad \forall\ t = 2, \ldots, n}$$

Scanned by CamScanner

Polynomial Fitting:

ex: $m_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \longrightarrow$ fit

minimize the sq. dist.

$$\rightarrow \sum_{t=1}^{n} (x_t - m_t)^2$$

Trend elimination by differencing:

estimate 'k' s.t $\nabla^k X_t \approx$ const. Then fit k-degree polynom.

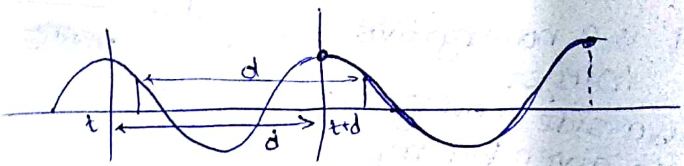
Estimation of Trend and Seasonality:

ex: Unemployement, removal of seasonal unemployement to observe true trend of unemployement

- let, $X_t = T_t + S_t + W_t$ (Additive Model)

with $E(W_t) = 0$

$$\boxed{S_{t+d} = S_t} \text{ and } \sum_{t=1}^{d} S_t = 0$$



$\nabla^k = (1-B)^k$

$\nabla_d = I - B^d$

$\nabla_d X_t = X_t - X_{t-d}$

$X_t = T_t + S_t + W_t$

$X_{t-d} = T_{t-d} + S_{t-d} + W_{t-d}$

$$\boxed{X_t - X_{t-d} = (T_t - T_{t-d}) + (W_t - W_{t-d})}$$

Criterion:
BIC
AIC

- Now $T_t - T_{t-d}$ can be removed by using above method.
- Hence S_t aswell can be calc.

Testing the Estimated Noise Sequence:

ex: Noise removal in Signal prediction

- If there is no dependence among these residuals, then we can regard them as observations of independent RVs & there is no further modelling to be done except to estimate their mean & variance.

→ The sample autocorrelation fn, 95%.

$$CI = \left(-1.96/\sqrt{n},\ +1.96/\sqrt{n}\right).$$

ACVF,

let $\{X_t\}$ be weakly stationary time series

ACVF of X_t is

$$\gamma_X(h) = \mathrm{cov}(X_{t+h}, X_t).$$

Autocorrelation of X_t,

$$\rho(X_t, X_{t+h})$$

$$= \frac{\gamma_X(h)}{\sqrt{V(X_{t+h})V(X_t)}} = \frac{\gamma_X(h)}{\sqrt{\gamma_X(0)\cdot\gamma_X(0)}} = \frac{\gamma_X(h)}{\gamma_X(0)}$$

ex. paired association

h=1

X_1
$X_2 \leftrightarrow X_1$ ⇒ (X_2, X_1)
$X_3 \leftrightarrow X_2$ (X_3, X_2)
⋮
$X_n \leftrightarrow X_{n-1}$ (X_n, X_{n-1})

∴ paired obs^{ns}.

$$\therefore \text{correlation} = \frac{(\Sigma x_i y_i) - (\bar{x}\bar{y})}{\sqrt{\sigma_x^2}\sqrt{\sigma_y^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\Sigma(x_i-\bar{x})^2\cdot\frac{1}{n}\Sigma(y_i-\bar{y})^2}}$$

$$\boxed{\hat{\rho}(h) \sim AN}$$

- Ljung and Box (1978) Test: (LB Test):

$$Q = n(n+2)\sum_{j=1}^{h}\hat{\rho}^2(j)/(n-j)$$ whose distrib^n is χ^2_h-distb^n (h → df)

- Non-parametric Test:

Rank test
Run test
Sign test.

If we don't know the distb^n of X, then non-param test. ***

$$\left.\begin{aligned} H_0 &: E(X) = \mu_0 \\ H_1 &: E(X) > \mu_0 \end{aligned}\right\}$$ Parametric Test.

"Rank Test:"

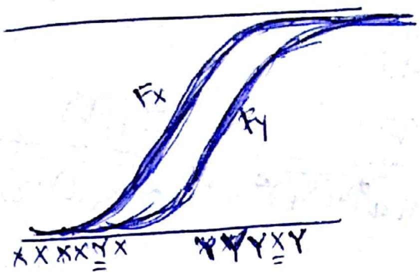
$x_1, x_2, x_3, \dots x_{n_1}$
$y_1, y_2, y_3 \dots y_{n_2}$

check if both come from same cdf

X is stochastically larger than Y?

⇓

⇒ say X > Y



① Arrange $(\underset{\sim}{X}, \underset{\sim}{Y})$ in inc. order.

if run length is less → well-mixed

a) run-length,
b) standardized-length, etc

Test of less pwr
∴ not preferred
⇒ "Run Test"

If it is > than prev. value ⇒ +ve
" " " < " " " ⇒ −ve

} Binomial Test

∴ series of + & − ⟹ "Sign Test"

→ Autocorrelation Fn (ACF):

Suppose that X_t is atleast weakly stationary T.S.

$E(X_t) = \mu$ and $\gamma_x(h) = E[(X_t - \mu)(X_{t+h} - \mu)]$.

Defn → $ACF(X_t, X_{t+h}) = \rho(h) = \dfrac{\gamma_x(h)}{\gamma_x(0)}$

~ Properties:

① $\rho(h) = \rho(-h) = \rho(|h|)$

② $R = ((\rho_{ij})) = ((\rho\,|i-j|))_{ij}$ $\quad 1 \le i,j \le n$

$\underset{\sim}{a}^T R \underset{\sim}{a} \geq 0 \Rightarrow R$ is a p.s.d. matrix

$\rho_{ij} = \dfrac{cov(X_i, X_j)}{\sqrt{V(X_i)V(X_j)}} = \dfrac{\sigma_{ij}}{\sigma_i \sigma_j}$

$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{pmatrix}$

$R_x = \begin{pmatrix} 1/\sigma_1 & & 0 \\ & 1/\sigma_2 & \\ 0 & & \ddots \; 1/\sigma_n \end{pmatrix} \Sigma_x \begin{pmatrix} 1/\sigma_1 & & 0 \\ & 1/\sigma_2 & \\ 0 & & \ddots \; 1/\sigma_n \end{pmatrix}$

Σ_x is also p.s.d. matrix $\Rightarrow$ due to which R is p.s.d. matrix

③ P.S.D. Function:

Defⁿ: A function f is said to be p.s.d. if,

$f : \cancel{\{1,\dots,n\}}\ \mathbb{Z} \to \mathbb{R}$ such that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} f(i-j)a_i a_j \geq 0 \quad \forall\, n \in \mathbb{N} \text{ and } \underset{\sim}{a} \neq \underset{\sim}{0}$$

$\Rightarrow$ Show that: $\gamma()$ & $\rho()$ are p.s.d. fns.

(i) To show:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \underbrace{\gamma(i-j)}_{\text{autocov. b/w } X_i \,\&\, X_j} a_i a_j \geq 0.$$

$$\Rightarrow \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(X_i, X_j)\, a_i a_j \geq 0$$

$$\Rightarrow \underset{\sim}{a}^T \Sigma_x \underset{\sim}{a} \geq 0$$

$$\Rightarrow \cancel{\mathrm{var}(\underset{\sim}{a}\underset{\sim}{a}) \geq 0}\ \mathrm{Var}(\underset{\sim}{a}^T \underset{\sim}{X}) \geq 0$$

(ii) since (i) σ_i, σ_j are +ve. $\therefore$ (ii) ≥ 0.

$\rightarrow$ Linear Process: (time dependent)

- An important class of weakly stationary times series:

- Defⁿ: $\boxed{X_t = \mu + \sum_{j=-\infty}^{+\infty} \psi_j W_{t-j}}$ $W_t \sim WN(0, \sigma_w^2)$ where, $\mu \in \mathbb{R}$

where $\mu \in \mathbb{R}$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, $W_j \sim WN(0, \sigma_w^2)$

[if $E|X_t| < \infty$]

(i) $E(X_t) = \mu$

Proof: $E(X_t) = E\left(\mu + \sum_{-\infty}^{\infty} \psi_j (W_{t-j})\right)$

$= \mu + \sum_{-\infty}^{\infty} \psi_j \cancel{E(W_{t-j})}^{\,0}$

$= \underline{\underline{\mu}}$

For (Z), if $\Sigma(|z| f(z)) < \infty$ then, $E(Z) = \Sigma z f(z)$

(ii) $\gamma_x(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-h} < \infty$

Proof: $Cov(X_t, X_{t+h}) = ?$

(i) Z

if $\sum |z| f(z)| < \infty$

then $E(z) = \sum z f(z)$

$$E|X_t| = ? \Rightarrow E\left|\mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}\right|$$

$$\leq E(|\mu|) + E\left|\sum_{j=-\infty}^{\infty} \psi_j W_{t-j}\right|$$

$$\leq \underbrace{E|\mu|}_{\text{finite}} + \sum_{j=-\infty}^{\infty} E|\psi_j| \underbrace{|W_{t-j}|}_{\text{finite?}}$$

$$\leq |\mu| + M \cdot \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

NOTE: R.T. $V(W) < \infty \Rightarrow E|W| < \infty$

$\because$ P.T. if $E(|X|^r) < \infty$

$\Rightarrow E(|X|^s) < \infty$ if $s \leq r$

↳ higher order moment is finite ⇒ lower order moment is finite

3 lines

$\therefore E|X_t| \leq$ finite + finite + finite $< \infty$

finite

General Result: *

consider a weakly stationary time series $\{X_t\}$ with zero mean, and define

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

where, $\mu \in \mathbb{R}$ and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ then,

(i) $E(Y_t) = \mu$

(ii) $\gamma_Y(h) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-h} \gamma_X(h-j+k)$, if exists.

$\{a_i\}$ $\quad \sum a_i < \infty \quad \Rightarrow$ conditionally convergent

but $\sum |a_i| \uparrow \infty \quad \Rightarrow$ absolute divergence

ex. $\infty \leftarrow = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$ divergent (absolute)

$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \cdots$ convergent (condition)

put even terms & odd terms together

$\Rightarrow \infty - \infty \Rightarrow \cdots ?$

$\therefore$ We first ensure $E|X_t| < \infty$ (*)

3/10/18

→ **General Result:**

$\{X_t\}$ weakly stationary process

$E(X_t) = 0 \quad V(X_t) < \infty$ (finite)

$$\boxed{Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}, \quad \text{where } \sum_{-\infty}^{\infty} |\psi_j| < \infty \quad \mu \in \mathbb{R}}$$

$\gamma_Y(h) = cov(Y_{t+h}, Y_t)$

$= cov(Y_{t+h} - \mu, Y_t - \mu)$

$= E\left(\sum_{-\infty}^{\infty} \psi_j X_{(t+h)-j} \cdot \sum_{k=-\infty}^{\infty} \psi_j X_{t-k}\right) - 0$

$\therefore E|X_t| \leq$ finite + finite · finite $< \infty$

finite

**** General Result: ***

Consider a weakly stationary time series $\{X_t\}$ with zero mean, and define

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

where $\mu \in \mathbb{R}$ and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ then,

(i) $E(Y_t) = \mu$

(ii) $\gamma_Y(h) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \gamma_X(h-j+k)$, if exists.

$\{a_i\}$ $\quad \sum a_i < \infty$ $\Rightarrow$ conditionally convergent

but $\sum |a_i| \uparrow \infty$ $\Rightarrow$ absolute divergence

ex: $\infty \leftarrow = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$ divergent (absolute)

$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \cdots$ convergent (conditional)

put even terms & odd terms together

$\Rightarrow \infty - \infty \Rightarrow \cdots ?$.

$\boxed{\therefore \text{ we first ensure } E|X_t| < \infty \ (\ast\ast)}$

3/10/18

→ General Result:

$\{X_t\}$ weakly stationary process

$E(X_t) = 0 \quad V(X_t) < \infty$ (finite)

$$\boxed{Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad \text{where } \sum_{-\infty}^{\infty} |\psi_j| < \infty \quad \mu \in \mathbb{R}}$$

$$\gamma_Y(h) = \text{cov}(Y_{t+h}, Y_t)$$

$$= \text{cov}(Y_{t+h} - \mu, Y_t - \mu)$$

$$= E\left(\sum_{j=-\infty}^{\infty} \psi_j X_{t+h-j} \cdot \sum_{k=-\infty} \psi_j X_{t-k} \right) - 0$$

Prereq check:

$\therefore$ check $E\left|\left(\sum \psi_j X_{t+h-j} \cdot \sum \psi_k X_{t-k}\right)\right| < \infty$

$$\Rightarrow E\left|\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} \psi_j \psi_k X_{t+h-j} X_{t-k}\right|$$

$$\leq E\left(\sum_{-\infty j}^{\infty}\sum_{-\infty k}^{\infty} |\psi_j||\psi_k| \, |X_{t+h-j} X_{t-k}|\right)$$

$$\leq \left(\sum_j \sum_k |\psi_j||\psi_k| \, E\left(|X_{t+h-j} X_{t-k}|\right)\right)$$

$$\leq M \cdot \sum_j \sum_k |\psi_j||\psi_k|$$

$$\leq M \cdot \left(\sum_j |\psi_j|\right)\left(\sum_k |\psi_k|\right).$$

VIMP

$\therefore \left|Corr\left(|X_{t+h-j}|, |X_{t-k}|\right)\right| \leq 1$

$Cov\left(|X_{t+h-j}|, |X_{t-k}|\right) \leq \sqrt{(V(|X_t|))^2}$

$V(|X_t|)$
$= E(|X_t|^2) - (E|X_t|)^2$
$\leq V(X_t) - (E|X_t|)^2$
$< \infty$

$$\therefore \; = E\left(\sum_{-\infty}^{\infty} \psi_j X_{t+h-j} \cdot \sum_{-\infty}^{\infty} \psi_k X_{t-k}\right)$$

$$= E\left(\sum_{-\infty j}^{\infty}\sum_{-\infty k}^{\infty} \psi_j \psi_k X_{t+h-j} X_{t-k}\right)$$

$$= \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} \psi_j \psi_k \, \cancel{E}^{\,Cov}\left(X_{t+h-j} X_{t-k}\right)$$

$$= \underline{\sum_{-\infty j}^{\infty}\sum_{-\infty k}^{\infty} \psi_j \psi_k \, \gamma_X(h-j+k)} \quad H.P.$$

~ An important class of W.S.T.S:

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j} \qquad W_t \sim WN(0, \sigma_w^2)$$

$$\gamma_X(h) = V_X(h) = \sum_j \sum_k \psi_j \psi_k \, \gamma_W(h-j+k)$$

[when lag (h) is non-zero
$\sigma_w^2 \; \cancel{\quad} = \gamma_X(h)$]

?

- $E(X_t) = \mu$
- $\gamma_x(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-h} < \infty$

~ Properties of WN:

if $j \neq k$, cov $= 0$
$j = k$, cov $= \sigma_w^2$

① WN is with mean zero & fixed finite variance & are uncorrelate

② WN need not be normally distributed

③ WN need not be iid

④ But iid sequence is always WN

⑤ WN is weakly stationary.

⑥ If a WN sequence is Normally distributed then it is strongly stationary.

→ AutoRegressive Process: (AR)

A T.S. $\{X_t\}$ is said to be an auto regressive process (of order one) if, $X_t = \phi X_{t-1} + W_t$, where $W_t \sim WN(0, \sigma_w^2)$

$\Rightarrow$ AR(1), if

$$W_t \sim WN(0, \sigma_w^2)$$

$$X_t = \phi X_{t-1} + W_t, \quad |\phi| < 1, \ \phi \neq 0$$

$$Y = \beta X + \epsilon$$ → except it occurs within the series.

$$\begin{aligned} X_t &= \phi X_{t-1} + W_t \\ &= \phi(\phi X_{t-2} + W_{t-1}) + W_t \\ &= \phi^2 X_{t-2} + (\phi W_{t-1} + W_t) \\ &\ \ \vdots \\ &= \phi^t X_0 + (\phi^{t-1} W_1 + \phi^{t-2} W_2 + \dots + W_t) \ (?) \\ &= \sum_{j=0}^{\infty} \phi^j W_{t-j}, \quad \phi \neq 0. \end{aligned}$$

→ trails backwards.

$\Rightarrow$ $\therefore$ AR(1) is a linear process:

① $\boxed{\psi_j = \begin{cases} \phi^j, & j \geq 0 \\ 0, & \text{o.w.} \end{cases}}$

② $\therefore$ $\boxed{E(X_t) = 0}$

③ $\boxed{\gamma_x(h) = \dfrac{\sigma_w^2 \phi^{|h|}}{1-\phi^2}}$, if $|\phi|<1$

4. $Var(X_t) = V\left(\sum_{j=0}^{\infty} \phi^j W_{t-j}\right)$

$= \sum_{j=0}^{\infty} \phi^{2j} * Var(W_{t-j})$

$= \sigma_w^2 \sum_{j=0}^{\infty} \phi^{2j}$

$= \dfrac{\sigma_w^2}{1-\phi^2} = \underline{\underline{\gamma_x(0)}}$ → $h=0$

$\gamma_x(h) = Cov(X_t, X_{t+h}) = cov\left(\sum_{j=-\infty}^{\infty} \phi^j W_{t-j}, \sum_{j=-\infty}^{\infty} \phi^j W_{t+h-j}\right)$

$= \cancel{Cov\left(\sum_{\infty} \phi^j W_{t-j}, \sum_{-\infty}^{\infty} \phi^j W_{t-j} + \sum\right.}$

$= Cov\left(X_t, \phi^h X_t + \sum_{j=0}^{h-1} \phi^j W_{t+h-j}\right)$

[∵ $X_{t+h-j} = W_{t+h} + \phi W_{t+h-1} + \cdots + \phi^h X_t$]

$= Cov(X_t, \phi^h X_t) + \cancel{Cov\left(X_t, \sum_{j=0}^{h-1} \phi^j W_{t+h-j}\right)}^{\,0}$

$= \dfrac{\sigma_w^2 \phi^{|h|}}{1-\phi^2}$

[∵ $\gamma_x(|h|) = \gamma_x(-h) = \gamma_x(h)$]

OR

$\gamma_x(h) = Cov(X_{t+h}, X_t)$

$= Cov(\phi X_{t+h-1} + W_{t+h}, X_t)$

$\gamma_x(h) = \phi \, Cov(X_{t+h-1}, X_t) = \phi \gamma_x(h-1)$

$= \phi^2 \gamma_x(h-2)$

$= \cdots \phi^h \gamma(0)$

$= \dfrac{\phi^h \sigma_w^2}{1-\phi^2}$

HW:
Strongly st?
OR NOT?
give examp

4/10/18

(4) If $\phi = 0$, then AR(1) process is a WN process.
(5) AR(1) is atleast a Weakly Stationary process.

AR(1) process is a linar process with $\mu = 0$ &

$$\psi_i = \begin{cases} 1 & , j=0 \\ \phi^j & , j \geq 1 \\ 0 & , \text{o.w.} \end{cases}$$

→ Ex: AR(1): ARIMA Integrated (p) Autoregressive (d) Moving average (q) $n=100$ $ar=0.7$

(Autocorrelation) ACF v.s. n

type=T

→ **Moving Average Process (MA):**

A T.S. $\{X_t\}$ is a MAP. (of order 1) if:

$$\boxed{X_t = W_t + \theta W_{t-1}}, \quad W_t \sim WN(0, \sigma_w^2).$$

$$X_t = 1 \cdot W_t + \theta W_{t-1} + 0 \cdot W_{t-2} + \cdots$$

(1) MA(1) Process is a linear process with $\mu = 0$

$$\psi_j = \begin{cases} 1 & , j=0 \\ \theta & , j=1 \end{cases}$$

(2) $\therefore E(X_t) = E(W_t + \theta W_{t-1})$
$= 0 + 0 = \underline{\underline{0}}$

(3) $\therefore \gamma_x(h) = \begin{cases} Var(X_t) = \sigma_w^2(1+\theta^2) & , \text{if } h=0 \\ Cov(X_t, X_{t-1}) = \theta\sigma_w^2 & , \text{if } h = \pm 1 \\ 0 & , \text{o.w.} \end{cases}$

$Cov(X_t, X_{t-1}) = cov(W_t + \theta W_{t-1}, W_{t-1} + \theta W_{t-2})$ ∵ on splitting till common terms, else $=0$.
$= \theta\sigma_w^2$

$Cov(X_t, X_{t-2}) = Cov(W_t + \theta W_{t-1}, W_{t-2} + \theta W_{t-3})$
$= \underline{\underline{0}}$

(4) MA(1) is a stationary process.

Str. Stationary?

① Check if X_t & X_{t+h} have same dist.

② (X_t, X_{t+1}) & (X_{t+h}, X_{t+h+1}) joint distbn's have same dist.

$$X_{t+h} = W_{t+h} + \theta W_{t+h-1}$$

$$E(X_{t+h}) = 0 + 0 = \underline{\underline{0}}$$

$$Var(X_{t+h}) = (1+\theta^2)\sigma$$?.

ex: $W_t = \begin{cases} Z \sim N(0,1), & \text{if } t = \text{even} \\ Y \sim exp(1) - 1, & \text{if } t = \text{odd} \end{cases}$

∴ Need not be strong Stationary.

✓ Better approach.

$$X_{10} \stackrel{d}{=} W_{10} + \theta W_9 \stackrel{d}{=} Z + \theta Y$$

$$X_{13} \stackrel{d}{=} W_{13} + \theta W_{12} \stackrel{d}{=} Y + \theta Z$$

∴ Not the same distbn.

∴ ALWAYS W.S.

Not always S.S.

• ACF ⇒ $\rho_x(h) = \dfrac{\gamma_x(h)}{\gamma_x(0)}$

ex. "AR with center zero"

∴ Use of sample ACF

T.S.	ACF
WN	zero
MA(q)	zero ; $\lvert h \rvert > q$
AR(p)	decays to zero exponentially

Scanned by CamScanner

b) Sample Autocovariance fn :

$$\text{cov}(X,Y) = E\left[(X-E(X))(Y-E(Y))\right].$$

$$\therefore \gamma(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_t-\bar{x})(x_{t+|h|}-\bar{x}) \quad \text{if } -n<h<n$$

scaled by 'n' NOT n-|h|' ?

for h = lag

h=2 (or h)

x_1
x_2

x_3	x_1
x_4	x_2
x_5	x_3
⋮	⋮
x_n	x_{n-2}

n-2 or (n-h) paired observations. for lag h.

c) Sample variance $\Rightarrow \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$?

d) ACF = $\hat{\rho}(h) = \dfrac{\gamma_X(h)}{\gamma_X(0)}$

→ **Sample Mean ($\bar{x}$):**

- $E(\bar{x}) = \mu$ (unbiased estimator)

** • $\text{Var}(\bar{x}) = \frac{1}{n}\left(\sum_{h=-n}^{n}\left[\left(1-\frac{|h|}{n}\right)\gamma(h)\right]\right) \longrightarrow 0$ as $n\to\infty$

- ~~Var~~ $\gamma(h) = \text{cov}(X_t, X_{t+h})$
  $= \text{cov}(X_{t-h}, X_t)$

  Assume $\sum_{-\infty}^{\infty}|\gamma(h)| < \infty$

$X_i \sim N(\mu, \sigma^2)$

- $\widehat{\text{Var}}(\bar{x}) = \frac{1}{n}\sum_{h=-n}^{n}\left(1-\frac{|h|}{n}\right)\hat{\gamma}(h)$ $\qquad \lim_{n\to\infty}\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \xrightarrow{d} N(0,1)$

- **Long Run variance:** → helps in standardization in long run for X_i's with diff. σ_i^2's

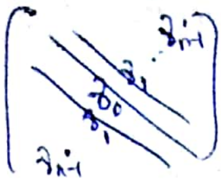
$$\lim_{n\to\infty}\sqrt{n}\,\text{Var}(\bar{x}) = \lim_{n\to\infty}\sum_{h=-\infty}^{\infty}\left(1-\frac{|h|}{n}\right)\gamma(h) = \sigma_w^2\left(\sum_{j=-\infty}^{\infty}\psi_j\right)^2$$

- $n\,\text{var}(\bar{x}) \longrightarrow \sum_{-\infty}^{\infty}\gamma(h)$, if $\sum_{-\infty}^{\infty}|\gamma(h)| < \infty$ as $n\to\infty$.

PROOF:

(ii) $\mathrm{Var}(\bar{x}) = \frac{1}{n} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h)$

$$V(\bar{x}) = \mathrm{Var}\left(\frac{\sum x_i}{n}\right) = \mathrm{Var}\left(\frac{\underset{\sim}{1}^T \underset{\sim}{x}}{n}\right)$$



$$= \frac{1}{n^2}\left(\underset{\sim}{1}^T \Sigma_x \underset{\sim}{1}\right)$$

$$= \frac{1}{n^2}\left[n\cdot\gamma(0) + 2(n-1)\cdot\gamma(1) + \dots 2\cdot(1)\cdot\gamma(n-1) + 2(0)\cdot\gamma(n)\right]$$

$$= \frac{1}{n^2} \sum_{h=-n}^{n}\left[\left(\frac{n-|h|}{n}\right)\gamma(h)\right]$$

$$= \frac{1}{n} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right)\gamma(h)$$

$\because \gamma(n-1) = \gamma(-(n-1))$

$\Downarrow$

$\{-n \dots -1, 0, 1, 2, \dots\}$

as $n \to \infty$,

$$\lim_{n\to\infty} \mathrm{Var}(\bar{x}) = \lim_{n\to\infty} \frac{1}{n} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right)\gamma(h)$$

$$\leq \lim_{n\to\infty} \frac{1}{n} \left| \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right)\gamma(h) \right|$$

$$\leq \lim_{n\to\infty} \frac{1}{n} \sum_{h=-n}^{n} \left|1 - \frac{|h|}{n}\right| |\gamma(h)|$$

$$\leq \lim_{n\to\infty} \frac{1}{n} \left(\sum_{h=-n}^{n} |\gamma(h)| \right) \downarrow 0 \quad (\text{as } n\uparrow\infty)$$

$\therefore \bar{x}$ is an unbiased estimator and consistent at ϕ

5/10/18

Weakly S. / Strongly S. $>$ mean & variance of (X_t) is free from 't'.

Bartlett's formula about $\hat{\rho}(h)$ (or $\hat{\rho}(h)$) rho hat h:

(general form of) (weakly stationary)

If $X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ with $E(X_t^4) < \infty$, then

approx. joint density of $(\hat{\rho}(i), \hat{\rho}(j))$ is:

$$\rightarrow \begin{pmatrix} \hat{\rho}(i) \\ \hat{\rho}(j) \end{pmatrix} \sim N\left[\begin{pmatrix} \rho(i) \\ \rho(j) \end{pmatrix}, \frac{1}{n}\begin{pmatrix} v_{ii} & v_{ij} \\ v_{ji} & v_{jj} \end{pmatrix}\right]$$

where,

$$v_{ij} = \sum_{h=1}^{\infty} \left(\rho(h+i) + \rho(h-i) - 2\rho(i)\rho(h)\right) \times \left(\rho(h+j) + \rho(h-j) \ldots\right.$$

$$\rightarrow \frac{1}{\gamma(0)} \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix}^{?} \text{ or } \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(n) \end{pmatrix} \sim N\left(\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(n) \end{pmatrix}, \frac{1}{n}\Gamma\right)$$

$$\hat{\rho}(h) = \frac{\gamma_x(h)}{\gamma_x(0)} \qquad v_{ij} = \sum_{h=1}^{\infty} (\quad) \times (\quad)$$

$\rightarrow$ $X_t \sim MA(1)$.

$X_t = W_t + \theta W_{t-1}$, $|\theta| < 1$

$\hat{\rho}(i) \sim N(\rho(i), \frac{1}{n} v_{ii})$ where $(i > 1)$

$$v_{ii} = \sum_{h=1}^{\infty} \left(\rho(h+i) + \rho(h-i) - 2\rho(h)\rho(i)\right)^2 \Rightarrow$$

if $h=0$, $\gamma_x(h) = \cancel{\sum_{h=1}^{\infty}}$ $\sigma_w^2(1+\theta^2)$

if $h=1$, $\gamma_x(h) = \sigma_w^2 \theta$

if $h=2$, $\gamma_x(h) = 0$

$\vdots$

$$\Rightarrow \frac{1}{\gamma_x(0)} \left[\sum_{h=1}^{\infty} \left(\gamma_x(h+i) + \gamma_x(h-i) - 2\gamma_x(h)\gamma_x(i)\right)^2\right]$$

$$\Rightarrow \frac{1}{\gamma_x(0)}$$

---

for $i=1$ $\Rightarrow$ $\overset{(h=0)}{\cancel{0}} + \overset{(h=1)}{(1-2\rho^2(1))} + \overset{(h=2)}{(\rho(1))^2} + 0 \ldots$

$= 1 - 3\rho^2(1) + 4\rho^4(1) \Rightarrow \left[\rho(1) = \frac{\theta}{1+\theta^2}\right]$

for i=2

$$v_{ii} = \sum_{h=1}^{\infty} \left(\rho(h+2)+\rho(h-2)-2\rho(h)\rho(2)\right)^2$$

$$\Rightarrow \underset{h=1}{(\rho(1))^2} + \underset{h=2}{(\rho(0))^2} + \underset{h=3}{(\rho(1))^2} + 0 + 0 \ldots$$

$\because \rho(-1)=\rho(1)$, $\rho(0)=1$

$$= 2\rho(1)^2+\rho(0)^2$$

$$= 1 + \frac{2\theta^2}{(1+\theta^2)^2}$$

$$v_{12} = \sum_{h=1}^{\infty} \left(\rho(h+1)+\rho(h-1)-2\rho(h)\rho(1)\right)\left(\rho(h+2)+\rho(h-2)-2\rho(h)\rho(2)\right)$$

i=1, j=2

$$= \underset{h=1}{(1-2\rho(1)^2)(\rho(1))} + \underset{h=2}{(\rho(1))(1)} + \underset{h=3}{(0)\cdot(\ldots)}$$

$$= 2\rho(1)\left[1-\rho(1)^2\right]$$

* **HW:** v_{ij} for AR(1)

AR(1) $\Rightarrow$ $\boxed{\hat{\rho}(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^{|h|}}$; $\boxed{X_t = W_t + \phi X_{t-1}}$; $|\phi|<1$

$$v_{11} \Rightarrow \underset{h=1}{(1-\phi^2)^2} + \underset{h=2}{\phi^2(1-\phi^2)^2} + \cdots = (1-\phi^2)^2(1+\phi^2+\phi^4+\cdots) = \frac{(1-\phi^2)^2}{1-\phi^2} = 1-\phi^2$$

$$v_{22} \Rightarrow \underset{h=1}{(\phi-\phi^3)^2} + \underset{h=2}{(1-\phi^4)^2} + \underset{h=3}{(\phi-\phi^5)^2} + \cdots$$

$$= \phi^2(1-\phi^2)^2 + (1-\phi^4)^2(1+\phi^2+\phi^4+\cdots)$$

$$= \frac{(1-\phi^2)^2(1+\phi^2)^2}{(1-\phi^2)} + (1-\phi^2)^2\phi^2 = (1-\phi^2)\left[1+\phi^4+2\phi^2+\phi^2-\phi^4\right]$$

$$= \underline{(1-\phi^2)(1+3\phi^2)}$$

$$v_{12} = \underset{h=1}{(1-\phi^2)\cdot(1-\phi^2)\phi} + \underset{h=2}{(1-\phi^2)\phi\cdot(1-\phi^4)} + \underset{h=3}{(1-\phi^4)\cdot\phi(1-\phi^4)} \cdots$$

$$= \phi(1-\phi^2)^2 + \phi(1+\phi^2)(1-\phi^2)^2 + \phi(1-\phi^4)^2\left[1+\phi+\phi^2+\cdots\right]$$

$$= 2\phi(1-\phi^2) + \phi(1-\phi^2)(1+\phi^2)^2(1+\phi)$$

$$= \underline{\phi(1-\phi^2)\left[2+(1+\phi)(1+\phi^2)^2\right]}$$

→ AR(p) and M...

Def^n: AR... as...

AR(p): ord...

Def^n:

** determined completely by the error.

AR(...

⇒ Conv...

weakness of notion inc.

AR(p) and MA(q) process:

Def^n: AR(p), an autoregressive process of $\boxed{\text{order } p}$ is defined as
$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = W_t.$$

AR(p): (order)
$$X_t = \phi X_{t-1} + W_t \longrightarrow AR(1) \quad |\phi|<1$$
$$X_t = \underbrace{\sum_{j=1}^{p} \phi_j X_{t-j}}_{\text{"p"-past time series values}} + \underbrace{W_t}_{\text{present error}}$$
"p"-past time series values and present error.

ex. AR(2)
$\Rightarrow X_t = \phi X_{t-2} + \phi X_{t-1} + W_t$

[$\because \phi_i = \phi^i, i>0$];

Def^n: MA(q), a moving average process of order q is defined as:
$$X_t = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q}$$

** determined completely by the error.

MA(q):
$$X_t = W_t + \theta X_{t-1}$$
$$= \underbrace{W_t}_{\text{present error}} + \underbrace{\sum_{j=1}^{q} \theta_j W_{t-j}}_{\text{past "q" errors}}$$
present error and past "q" errors.

AR(1): $X_t = \phi X_{t-1} + W_t$, $|\phi|<1$, $\phi \neq 0$
then AR(1) process is equivalent to MA(∞) process

$$\underset{AR(1)}{X_t} \overset{d}{=} \underset{MA(\infty)}{Y_t}$$
X_t & Y_t have same CDF

$\Rightarrow$ Converges in mean square:

seq. of RV $Y_1, Y_2, \dots$ converges in mean square to Z if for which: $\boxed{\lim_{n\to\infty} E(Y_n - Z)^2 = 0}$

convergence in mean square $\Rightarrow$ convergence in probability $\Rightarrow$ convergence in distb^n (BUT NOT THE OTHER WAY AROUND in GENERAL)

Almost sure convergence or wp1

weakness of notion inc.

$\therefore \Rightarrow \lim_{n\to\infty} E(Y_n - Z)^2 \to 0$

$\Rightarrow \lim_{n\to\infty} P(|Y_n - Z| > \epsilon) \to 0$

$\Rightarrow F_{Y_n}(t) \longrightarrow F_Z(t)$, $\forall$ t where F_Z is continuous.

Consider, AR(1) & MA(∞)

$$\Rightarrow \lim_{n\uparrow\infty} E\left(X_n - \sum_{j=0}^{n-1} \phi^j W_j\right)^2 = 0$$

AR(1) $\Rightarrow X_t = \phi X_{t-1}$ ...
$W_t \sim WN(0, ...$
$|\phi| < 1$

$X_t \sim AR(1)$

$$X_n = W_n + \phi X_{n-1}$$
$$= W_n + \phi W_{n-1} + \phi^2 W_{n-2} + \cdots$$
$$\vdots$$
$$= W_n + \phi W_{n-1} + \cdots + \phi^{n-(k+1)} W_{n-(k+1)}$$

8/10/18

→ **Mean Square Convergence**

series of RVs
$X_1, X_2, \ldots$ converges in M.S. to Z if:

If X is another RV such that,

$$\boxed{\lim_{n\to\infty} E(X_n - X)^2 = 0}$$
$$\Downarrow$$
$F_{X_n}(t) \longrightarrow F_X(t)$ where $F_X(t)$ is continuous.

$X_n \xrightarrow{L_2} X$ (Mean squared converges)

$$\Downarrow$$

$X_n \xrightarrow{d} X$ (convergence in distbn)

- consider AR(1) process $X_t = \phi X_{t-1} + W_t$, where $W_t \sim WN(0, \sigma_w^2)$ and $|\phi| < 1$

$$\Rightarrow X_t = W_t + \phi X_{t-1}$$
$$= W_t + \phi(W_{t-1}) + \phi^2 X_{t-2}$$
$$= W_t + \phi W_{t-1} + \phi^2 W_{t-2} + \phi^3 X_{t-3}$$
$$= \left(\sum_{j=0}^{k} \phi^j X_{t-j}\right) + \phi^{k+1} X_{t-(k+1)}$$

(assuming $k\uparrow\infty$)

$$E\Big(\underbrace{X_t}_{\sim AR(1)} - \underbrace{\sum_{j=0}^{k} \phi^j X_{t-j}}_{MA(k)}\Big)^2 = E\big(\phi^{2(k+1)} \cdot \underbrace{X^2_{t-(k+1)}}_{\text{bounded.}}\big)$$

VIMP for exam

$$= \phi^{2(k+1)} \cdot Var(X_{t-(k+1)})$$
$$= \phi^{2(k+1)} \cdot \sigma_w^2$$
$$\longrightarrow 0 \text{ (if } k\uparrow\infty)$$

$\left[\because |\phi| < 1 \quad \therefore |\phi|^\infty \to 0\right]$

As $\{X_t\}$ is an AR process with $|\phi|<1, \Rightarrow E(X_{t-k-1}^2)<\infty$

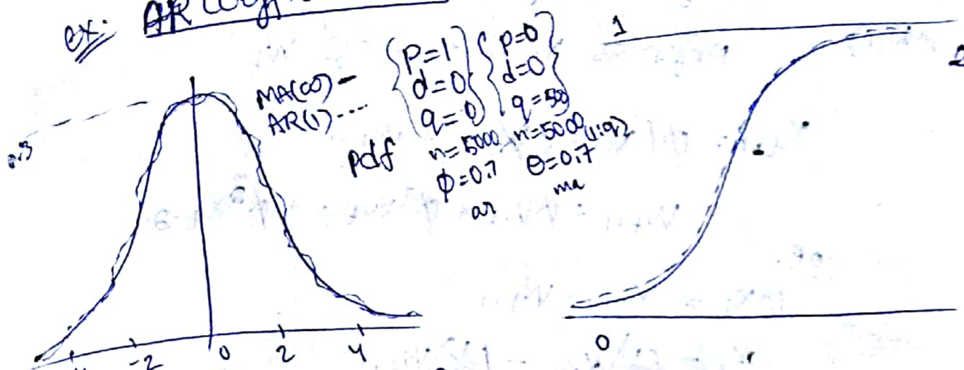
$$\therefore \lim_{k\uparrow\infty} E\left(X_t - \sum_{j=0}^{k}\phi^j X_{t-j}\right)^2$$

$$= \lim_{k\uparrow\infty} \phi^{2k+2} \cdot E(X_{t-k-1}^2) \longrightarrow 0 \quad \therefore MA(k) \xrightarrow{L_2} AR(1) \text{ as } k\uparrow\infty$$

$\therefore$ Convergence in M.S. $\Rightarrow$ convergence in Probability $\Rightarrow$ convergence in dist.^{tn} BUT NOT OTHER WAY AROUND IN GENERAL

As $MA(K)$ converges to $AR(1)$ process, it implies that $AR(1)$ process converges to distribution as $K\uparrow\infty$.

ex: ~~AR~~ AR coefficient $\Rightarrow$ "ϕ" $=0.7$



N=5000

Bandwidth $\Rightarrow$ $\underbrace{[\mu-\sigma, \mu+\sigma]}_{2\sigma=0.2316}$

→ **Casuality:**

- A linear process X_t is a casual fn. of W_t if,

$$\boxed{X_t = \left(1+\sum_{i=1}^{\infty}\phi_i B^i\right) W_t} \equiv X_t = \left(1+\sum_{i=1}^{\infty}\phi_i B^i\right) W_t$$

where, $\sum_{i=1}^{\infty}|\phi_i|<\infty$

$AR(1)$ i.e. $X_t = \phi X_{t-1} + W_t$ is casual, iff. $|\phi|<1$

i.e. $1-\phi z$ has a soln out of the unit circle on complex plane $\mathbb{C}$.

$\Rightarrow \boxed{z=\phi^{-1}}$ has soln

$\phi = x+iy$, $x,y<1$

$\phi^{-1} = \dfrac{1}{x+iy} = \dfrac{x-iy}{x^2+y^2}$

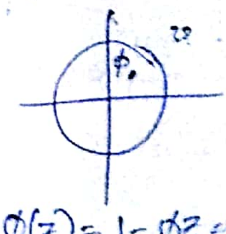
$\sqrt{(x^2+y^2)}<1$

$\therefore \dfrac{1}{x^2+y^2}>1$

$\therefore |\phi^{-1}|>1$

$x>x^2$

$|\phi|<1$



$\phi(z) = 1-\phi z = 0$

$z = 1/\phi$

$|z|>1$, $|\phi|<1$

Scanned by CamScanner

Invertibility:

- A linear process x_t is ~~irrever~~ invertible fn of w_t, if the

$$\boxed{w_t = \left(1+\sum_{i=1}^{\infty} \theta_i B^i\right) x_t}, \quad \text{where } \sum_{i=1}^{\infty} |\theta_i| < \infty$$

ex MA(1) i.e. $x_t = w_t + \theta w_{t-1}$, is invertible, iff $|\theta|<1$, i.e. $1+\theta z$ has a soln out of unit circle on complex plane $\mathbb{C}$

- $|\phi|>1$ $x_{t+1} = \phi x_t + w_{t+1}$

AR(1)   Express x_t in terms of w_t

$$x_{t+1} = \phi(w_t + \phi x_{t-1}) + w_{t+1}$$
$$= w_{t+1} + \phi w_t + \phi^2 w_{t-2} + \phi^3 x_{t-2}$$

OR

$$\phi x_t = x_{t+1} - w_{t+1}$$
$$x_t = \left(\frac{1}{\phi}\right) x_{t+1} - \left(\frac{1}{\phi}\right) w_{t+1}$$
$$\cancel{= \frac{1}{\phi}(\phi x_t + w_{t+1}) - \frac{1}{\phi}(w_{t+1})}$$
$$= (\phi^{-1}) x_{t+1} - (\phi^{-1}) w_{t+1}$$
$$= -(\phi^{-1}) w_{t+1} - (\phi^{-2}) w_{t+2} - (\phi^{-3}) w_{t+3} \ldots$$

$$\sum_{j=1}^{\infty} |\phi^{-j}| < \infty \quad \therefore \text{Non-casual}$$

"Random Walk"
$x_t = x_{t-1} + w_t$

$|1/\phi| < 1$

all the error terms are coming from the future
∴ Non-casual.

AR-MA (p,q) (Auto-Regressive – Moving Average Process)

An ARMA (p,q) process, x_t is a stationary process

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} \cdots - \phi_p x_{t-p} \quad \text{(order 'p')}$$
$$= w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \quad \text{(order 'q')}$$

"p+1" components   "q+1" components.

where, $w_i \sim WN(0, \sigma_w^2)$

$\therefore$ AR(p) = ARMA(p, 0)
MA(q) = ARMA(0, q)

ARMA(1,1)

AR coefficient ϕ
MA coefficient Θ $\Rightarrow$ $\boxed{X_t - \phi X_{t-1} = W_t + \Theta W_{t-1}}$

HW: What is the conditions on Θ & ϕ such that: * Endsemester

(i) ARMA(1,1) is invertible?

(ii) ARMA(1,1) is casual?

(iii) ARMA(1,1) is stationary?

?

ARMA(1,1) $\Rightarrow$ $X_t = W_t + (\Theta+\phi)\sum_{j=1}^{\infty}\phi^{j-1}W_{t-j}$

$\therefore$ (ii) casual ~~Invertible~~ $\Rightarrow$ $|(\Theta+\phi)\phi^{j-1}| < 1 \quad \forall j>0$

$\Rightarrow |(\Theta+\phi)\sum_{j=1}^{\infty}\phi^{j-1}| < \infty$

Invertible:

(i) $X_t - W_t - \Theta W_{t-1} = \phi X_{t-1}$

$X_{t-1} = \phi^{-1}(X_t - W_t - \Theta W_{t-1})$

$[\because X_t = \phi^{-1}(X_{t+1} - W_{t+1} - \Theta W_t)$

$= \cancel{\phi^{-1}}$
$\phi^{-2}X_{t+1} - \phi^{-2}W_{t+1} - (\Theta\phi^{-1} + 1)\phi^{-1}W_t - \phi^{-1}\Theta W_{t-1}$

$= -\phi^{-1}\Theta W_{t-1} - (\Theta\phi^{-1}+1)\phi^{-1}W_t - (\Theta\phi^{-1}+1)\phi^{-2}W_{t+1} \cdots$

$= \left|-[(\Theta\phi^{-1}+1)]\sum_{i=1}^{\infty}\phi^{-i}\right| < \infty$

(iii) Stationary a) $E(X_t) = \text{const.}$?

$= E(W_t + (\Theta+\phi)\sum_{j=1}^{\infty}\phi^{j-1}W_{t-j}) = \underline{0}$

prereq. $|(\Theta+\phi)|\sum_{j=1}^{\infty}|\phi|^{j-1}E(W_{t-j})| < \infty$

$|\Theta+\phi|\sum_{j=1}^{\infty}|\phi|^{j-1} < \infty$

b) $Var(X_t) = \text{const.}$? prereq.

$= E\left(\left|W_t^2 + \underbrace{2W_t(\Theta+\phi)\sum_{j=1}^{\infty}\phi^{j-1}W_{t-j}} + (\Theta+\phi)^2\sum_{j=1}^{\infty}\phi^{2(j-1)}W_{t-j}^2\right|\right)$

$= \sigma_W^2\left[1 + (\Theta+\phi)^2\sum_{j=1}^{\infty}\phi^{2(j-1)}\right] + \left|2(\Theta+\phi)\sum_{j=1}^{\infty}\left(\phi^{j-1}\cancel{E(W_t W_{t-j})}^{0}\right)\right|$

$= \sigma_W^2\left[1 + (\Theta+\phi)^2\sum_{j=1}^{\infty}\phi^{(2(j-1))}\right] < \infty$

$= (\Theta+\phi)^2\sum_{j=1}^{\infty}|\phi|^{2(j-1)} < \infty$

→ ARMA(p,q)

- Backshift operator B, such that $BX_t = X_{t-1}$.
- $\Phi_p(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$, if $\phi_p \neq 0$
- $\Theta_q(z) = 1 + \sum_{i=1}^{q} \theta_i z^i$, if $\theta_q \neq 0$

then ARMA(p,q) model can be written as

$$\Phi_p(B)X_t = \Theta_q(B)W_t$$

* if $\Phi_p(z)$ and $\Theta_q(z)$ have no common factor i.e. not a lower order ⇒ ARMA model is possible

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right)X_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right)W_t$$

→ ARMA(1,1)

$$X_t - \phi X_{t-1} = W_t + \theta W_{t-1}$$

$$(1-\phi B)X_t = (1+\theta B)W_t$$

$$X_t = (1-\phi B)^{-1}(1+\theta B)W_t$$

$$= (1+\phi B+\phi^2B^2+\phi^3B^3+\cdots)(1+\theta B)W_t$$

$$= \left[1 + B(\theta+\phi) + B^2(\theta+\phi)\phi + B^3(\theta+\phi^2)\phi + \cdots\right]W_t$$

?

$$\boxed{X_t = \sum_{j=0}^{\infty} \psi_j W_j}$$

$$\Rightarrow \psi_j = \begin{cases} j=t, & 1 \\ j=t-1, & (\theta^{t-j}+\phi^{t-j}) \\ j\le t-2, & (\theta^{t-j}+\phi^{t-j})\phi \end{cases}$$

$$\therefore \boxed{X_t = W_t + (\theta+\phi)\sum_{j=1}^{\infty}\phi^{j-1}W_{t-j}}$$

$$\psi_j = \begin{cases} 1, & j=0 \\ (\theta+\phi)\phi^{j-1}, & j\neq 0 \text{ or } j>0 \end{cases}$$

$$\gamma(0) = \sum_{j=1}^{\infty}\psi_j^2\sigma_w^2 = \sigma_w^2\left(1+\left[\psi_1^2+\psi_2^2+\cdots\right]\right)$$

$$= \sigma_w^2\left[(\theta+\phi)^2\left[\underbrace{\phi^{2(0)}}_{=1}+\phi^2+\phi^4+\cdots\right]+1\right]$$

$$= \sigma_w^2\left[\frac{(\theta+\phi)^2}{1-\phi^2}+1\right] = \sigma_w^2\left[\frac{\theta^2+2\theta\phi+1}{1-\phi^2}\right]$$

$$\therefore \gamma(h) = \begin{cases} \sigma^2\left[1 + \frac{(\theta+\phi)^2}{1-\phi^2}\right], & \text{if } h=0 \\ \sigma^2\left[(\theta+\phi) + \frac{\phi(\theta+\phi)^2}{1-\phi^2}\right], & \text{if } h=1 \\ \gamma(1)\cdot\phi^{h-1}, & \text{if } h\geq 2 \end{cases} \quad \& \quad \boxed{E(X_t)=0}$$

$$x_t - \phi x_{t-1} = w_t + \theta w_{t-1}, \text{ with } |\phi|<1$$

$$\gamma(1) = \cancel{\sigma^2[\theta+\phi+\phi} \; \sigma_w^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1}$$

$$= \sigma_w^2\left[(\theta+\phi) + (\theta+\phi)^2 \sum_{j=0}^{\infty} \phi^{(i+j-2)}\right]$$

$$= \sigma_w^2\left[(\theta+\phi) + \frac{(\theta+\phi)^2\phi}{1-\phi^2}\right]$$

$$\gamma(h) = \sigma_w^2\left[(\theta+\phi)\phi^{h-1} + (\theta+\phi)^2 \sum_{i=1}^{\infty} \phi^{2i+h-2}\right]$$

$$= \sigma_w^2\left[(\theta+\phi)\phi^{h-1} + \frac{(\theta+\phi)^2\phi^h}{1-\phi^2}\right]$$

$$= \phi^{h-1}\cdot\sigma_w^2\left[(\theta+\phi) + \frac{(\theta+\phi)^2\phi}{1-\phi^2}\right]$$

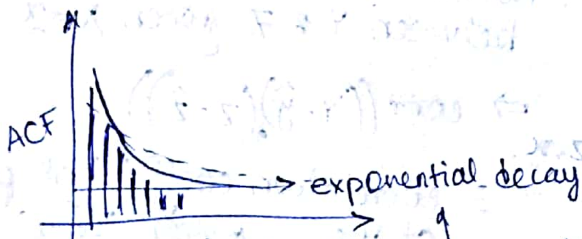
$$= \underline{\underline{\phi^{h-1}\gamma(1)}}$$

10/10/18

AR(1)

$\gamma_x(h) \propto \phi^{|h|}$

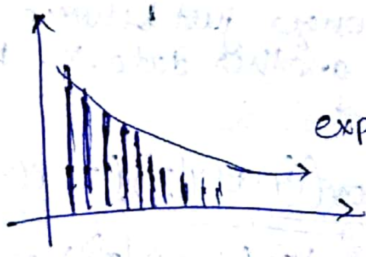
~~$\gamma_x(0) = const.$~~



BOTH

ARMA(1,1)

$\gamma_x(h) \propto \phi^{|h|}$



exponential decay

PACF is used to classify these two

MA(1) ✗ exponential

→ Partial Autocorrelation Coefficient (PACF)

To classify AR, MA, ARMA. (& to indicate order to some e

- **Partial correlation coefficient:**

$$\begin{pmatrix} Y \\ Z \\ \underset{\sim}{X} \end{pmatrix} \Rightarrow \quad \begin{array}{l} E(Y) = \mu_y \\ E(Z) = \mu_z \\ E(\underset{\sim}{X}) = \underset{\sim}{\mu}_x \end{array}$$

$$\Sigma = \left(\begin{array}{cc|cccc} \sigma_{yy} & \sigma_{yz} & \sigma_{yx_1} & \sigma_{yx_2} & \cdots & \sigma_{yx_p} \\ \sigma_{zy} & \sigma_{zz} & \sigma_{zx_1} & \sigma_{zx_2} & \cdots & \sigma_{zx_p} \\ \hline \sigma_{yx_1} & \sigma_{zx_1} & & & & \\ \sigma_{yx_2} & \sigma_{zx_2} & & \Sigma_x & & \\ \vdots & \vdots & & & & \\ \sigma_{yx_p} & \sigma_{zx_p} & & & & \end{array} \right)$$

Σ_{yz}

$\Sigma^T_{yz \cdot x}$ $2 \times p$

predict,

Y by $E(Y|\underset{\sim}{X} = \underset{\sim}{x})$ $\qquad$ Y & Z → known

& Z by $E(Z|\underset{\sim}{X} = \underset{\sim}{x})$ $\qquad$ $\underset{\sim}{X}$ → given

$\therefore \hat{Y} = E(Y|\underset{\sim}{X} = \underset{\sim}{x})$

$\hat{Z} = E(Z|\underset{\sim}{X} = \underset{\sim}{x})$

∴ Partial correlation coefficient:
between Y & Z given $X = \underset{\sim}{x}$

(rho)

$$\rho_{yz,\underset{\sim}{x}} \Rightarrow corr\left((Y-\hat{y}),(z-\hat{z})\right)$$

= correlation coefficient prediction error of Y & Z, given $X (= \underset{\sim}{x})$

$y, z \Rightarrow$ assume as scalars

* It is no longer just between Y & Z, it is predicted given another data, X. Hence the "partial".

$$\boxed{\rho_{yz,\underset{\sim}{x}} = \frac{cov\left((Y-E(Y|\underset{\sim}{x})),(Z-E(Z|\underset{\sim}{x}))\right)}{\sqrt{Var(Y-E(Y|\underset{\sim}{x}))\,Var(Z-E(Z|\underset{\sim}{x}))}}} \longrightarrow (i)$$

If $\begin{pmatrix} Y \\ Z \\ \underset{\sim}{X} \end{pmatrix}$ are normally distributed, then the conditional distributi' of $\begin{pmatrix} Y \\ Z \end{pmatrix}$ given $X = \underset{\sim}{x}$ is :

$$\left(\begin{matrix} Y \\ Z \end{matrix} \middle| \underset{\sim}{X} = \underset{\sim}{x} \right) \sim N\left(\left(\underset{(2\times1)}{\begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix}} + \underbrace{\underset{(2\times p)}{\Sigma^T_{yz\cdot x}} \cdot \underset{(p\times p)}{\Sigma^{-1}_x} \cdot \underset{(p\times 1)}{(\underset{\sim}{x} - \mu_x)}}_{(2\times1)} \right), \; \underset{(2\times2)}{\Sigma_{yz}} - \underset{(2\times p)}{\Sigma^T_{yz\cdot x}} \underset{(p\times p)}{\Sigma^{-1}_x} \underset{(p\times2)}{\Sigma_{yz\cdot x}} \right)$$

Dist^bn of Predicted Y, Z

mean is all known

Same!

NOTE:
* Recall : Bivariate Normal Dist^bn

$(\omega_1, \omega_2) \propto B.N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho_{12})$

$\therefore E(Y|x) = \cdots$

$$E(\omega_2|\omega_1) = \mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(\omega_1 - \mu_1)$$

$$\text{cov}(\omega_1, \omega_2) = \rho\sigma_1\sigma_2$$

$$\underbrace{\begin{pmatrix} Y \\ Z \end{pmatrix} - \left[\begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix} + \Sigma^T_{yz\cdot \underset{\sim}{x}} \cdot \Sigma^{-1}_x \left(\underset{\sim}{x} - \underset{\sim}{\mu} \right) \right]}_{\text{error}} \sim N\left(\underset{\sim}{0}, \Sigma_{yz} - \Sigma^T_{yz\cdot \underset{\sim}{x}} \Sigma^{-1}_X \Sigma_{yz\cdot \underset{\sim}{x}} \right)$$

Now, for variance of error:

$$\begin{pmatrix} \sigma_{yy\cdot \underset{\sim}{x}} & \sigma_{yz\cdot \underset{\sim}{x}} \\ \sigma_{zy\cdot \underset{\sim}{x}} & \sigma_{yy\cdot \underset{\sim}{x}} \end{pmatrix} = ?$$

wkt. $\rho_{yz\cdot \underset{\sim}{x}} = \cdots$ (i)

$$= \frac{\sigma_{yz\cdot x}}{\sqrt{\sigma_{yy\cdot \underset{\sim}{x}} \cdot \sigma_{zz\cdot \underset{\sim}{x}}}}$$

wkt.

$$\left(\begin{array}{cc|c} \sigma_{yy} & \sigma_{yz} & \sigma^T_{y\underset{\sim}{x}} \\ \sigma_{zy} & \sigma_{zz} & \sigma^T_{z\underset{\sim}{x}} \\ \hline \sigma_{y\underset{\sim}{x}} & \sigma_{z\underset{\sim}{x}} & \Sigma_x \end{array} \right)$$

$$\boxed{\rho_{yz\cdot \underset{\sim}{x}} = \frac{\left(\sigma_{yz} - \sigma^T_{y\underset{\sim}{x}} \Sigma^{-1}_x \sigma_{z\underset{\sim}{x}} \right)}{\sqrt{\left(\sigma_{yy} - \sigma^T_{y\underset{\sim}{x}} \Sigma^{-1}_x \sigma_{y\underset{\sim}{x}} \right)\left(\sigma_{zz} - \sigma^T_{z\underset{\sim}{x}} \Sigma^{-1}_x \sigma_{z\underset{\sim}{x}} \right)}}}$$

Now, what is PACF?

11/10/18

PACF:

$h=1, \quad \alpha(1) = PACF(1) = \rho(1) = \frac{\gamma(1)}{\gamma(0)}$

$h \geq 2, \alpha(h) = ?$

X_0	$X_1 \cdots X_{h-1}$	X_h
Y	$\underset{\sim}{X_{h-1}}$	Z

Corr. btwn prediction
$X_h - E(X_h | X_1, \cdots X_{h-1})$ &
$X_h - E(X_0 | X_1 \cdots X_{h-1})$.

Partial Autocorrelation function $\alpha(h)$, between X_h & X_0 for given $X_1, \cdots X_{h-1}$, the correlation btwn the prediction errors:

$$X_h - E(X_h | X_1, \cdots X_{h-1}), \quad X_0 - E(X_0 | X_1, \cdots X_{h-1})$$

$$\alpha(h) = \frac{\gamma(h) - \tilde{\gamma}^T_{h-1}(1) \Gamma^{-1}_{h-1} \gamma_{h-1}(1)}{\gamma(0) - \gamma^T_{h-1}(1) \Gamma^{-1}_{h-1} \gamma_{h-1}(1)}$$

where, $\tilde{\gamma}^T_{h-1}(1) = (\gamma(h-1), \cdots \gamma(1))^T$

and, $\gamma^T_{h-1}(1) = (\gamma(1), \gamma(2) \cdots \gamma(h-1))^T$

and, $\Gamma_{h-1} = ((\gamma(|i-j|)))_{ij}$

$\mu_0 = E(X_0) = 0$

$\mu_h = E(X_h) = 0$

$\mu_i = E(X_i) = 0$

$i = 1, 2, \cdots h-1$

		X_0	X_1	X_2	$\cdots$	X_{h-1}	X_h
Y	X_0	$\gamma(0)$	$\gamma(1)$	$\gamma(2)$	$\cdots$	$\gamma(h-1)$	$\gamma(h)$
$\underset{\sim}{X}$	X_1	$\gamma(1)$	$\gamma(0)$	$\gamma(1)$	$\cdots$	$\gamma(h-2)$	$\gamma(h-1)$
	X_2	$\gamma(2)$	$\gamma(1)$	$\gamma(0)$	$\cdots$	$\gamma(h-3)$	$\gamma(h-2)$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	X_{h-1}	$\gamma(h-1)$	$\gamma(h-2)$	$\cdots$		$\gamma(0)$	$\gamma(1)$
Z	X_h	$\gamma(h)$	$\gamma(h-1)$	$\cdots$		$\gamma(1)$	$\gamma(0)$

Γ_{h-1}

$\therefore \Sigma_X = \Gamma_{h-1}$

$\sigma_{Y\underset{\sim}{X}} = \gamma_{h-1}(1)$

$\sigma_{Z\underset{\sim}{X}} = \tilde{\gamma}_{h-1}(1)$

$\sigma_{YY} = \sigma_{ZZ} = \gamma(0)$

$\sigma_{YZ} = \gamma(h)$

$$\therefore \rho_{YZ\cdot X} = \frac{\sigma_{YZ} - \sigma_{YX}^T \cdot \Sigma_X^{-1}\sigma_{ZX}}{\sqrt{(\sigma_{YY}-\sigma_{YX}^T\Sigma_X^{-1}\sigma_{YX})(\sigma_{ZZ}-\sigma_{ZX}^T\Sigma_X^{-1}\sigma_{ZX})}}$$

$\therefore$ PACF, "α" = ?

$$\alpha(h) = \frac{\gamma(h) - \gamma_{h-1}^T(1)\,\Gamma_{h-1}^{-1}\,\tilde{\gamma}_{h-1}(1)}{\underbrace{\gamma(0) - \gamma_{h-1}^T(1)\,\Gamma_{h-1}^{-1}\,\gamma_{h-1}(1)}}$$

$$\sqrt{\left(\gamma(0) - \gamma_{h-1}^T(1)\,\Gamma_{h-1}^{-1}\,\gamma_{h-1}(1)\right)\left(\gamma(0) - \tilde{\gamma}_{h-1}^T(1)\,\Gamma_{h-1}^{-1}\,\tilde{\gamma}_{h-1}(1)\right)}$$

NOTE: ex. $(1\ 2\ 3) \rightarrow (2\ 3\ 1)$

$$\underset{3\times3}{\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}} \underset{3\times1}{\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}} = \underset{3\times1}{\begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}}$$

$(1,2,3) \rightarrow (3,2,1)$

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

$\downarrow$ R_3. (Reverse Matrix)

$$\Rightarrow \underbrace{\tilde{\gamma}_{h-1}^T(1)}\,\Gamma_{h-1}^{-1}\,\underbrace{\tilde{\gamma}_{h-1}(1)}$$

$$\Rightarrow (R_3\gamma_{h-1}(1))^T\,\Gamma_{h-1}^{-1}\,(R_3\gamma_{h-1}(1)) \Rightarrow \gamma_{h-1}^T(1)\underbrace{(R_3^T\Gamma_{h-1}^{-1}R_3)}\gamma_{h-1}(1)$$

$$\cancel{\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}} =$$

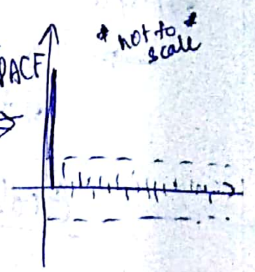
$\Rightarrow \boxed{\Gamma_{h-1}^{-1}}$

only because Γ_{h-1}^{-1} is symmetric else diff.

---

classification!

Model	ACF	PACF
AR(p)	decays	zero $(h>q)$
MA(q)	0 $(h>p)$	~~zero (h>q)~~ decays
ARMA(p,q)	decays	~~zero~~ decays



$\rightarrow$

$$\begin{pmatrix} Y \\ \underset{\sim}{X} \end{pmatrix} \sim \left(\begin{pmatrix} \mu_Y \\ \underset{\sim}{\mu_X} \end{pmatrix}, \begin{pmatrix} \sigma_{YY} & \underset{\sim}{\sigma}_{YX}^T \\ \underset{\sim}{\sigma}_{YX} & \Sigma_X \end{pmatrix} \right)$$ $\leftarrow$ assuming all three are KNOWN

Predict Y as a linear function of $\underset{\sim}{X}$

$$\hat{y} = (\underset{\sim}{\beta}^T_{1\times n} \underset{\sim}{X}_{n\times 1})$$

Least Square Condition is:

$$S = E[(Y - \beta^T X)^2]$$ $\rightarrow$ theoretical expectation (other form of the empirical expectation) is minimum ~~for~~ wrt. $\underset{\sim}{\beta}$.

W.L.G. Assume $E(Y) = \mu_Y = 0$, $E(\underset{\sim}{X}) = \underset{\sim}{\mu_X} = \underset{\sim}{0}$

$$\frac{\partial S}{\partial \underset{\sim}{\beta}} = 0 \Rightarrow ?$$

$$E(Y\underset{\sim}{X}) = E(XX^T)\underset{\sim}{\beta}$$

$$\Rightarrow E(Y\underset{\sim}{X}) = \begin{pmatrix} E(Y\cdot X_1) \\ E(Y\cdot X_2) \\ \vdots \\ E(Y\cdot X_n) \end{pmatrix}_{(n\times 1)} = \underset{\sim}{\sigma}_{Y\underset{\sim}{X}\,(n\times 1)}$$

$$\Rightarrow E(XX^T)\underset{\sim}{\beta} \Rightarrow \underset{\sim}{\sigma}_{Y\underset{\sim}{X}} = \Sigma_X \underset{\sim}{\beta}$$

$$\therefore \boxed{\underset{(n\times 1)}{\hat{\underset{\sim}{\beta}}} = \underset{(n\times n)}{\Sigma_X^{-1}} \cdot \underset{(n\times 1)}{\underset{\sim}{\sigma}_{Y\underset{\sim}{X}}}}$$

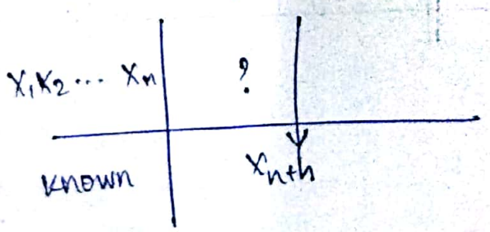
HW: Is it possible to have more than one estimate of $\underset{\sim}{\beta}$ satisfying LS condition? NO! $\because$ Only one BLUE

22/10/18

$\rightarrow$ Consider the time series $\{X_t\}$ with mean factor and autocovariance matrix completely known.

Goal: Estimate X_{n+h}, $h \geq 1$

when $\{X_n, X_{n-1}, \ldots, X_1\}$ given



If $\{X_t\}$ is stationary then:

① $\hat{X}_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n-i+1} - \mu)$ is a linear estimator of X_{n+h}

② Defining $\underset{\sim}{a} = (a_1 \dots a_n)$ as regression coefficients

$\hat{\underset{\sim}{a}} = \Gamma_n^{-1} \underset{\sim}{\gamma}_n(h)$

$$\begin{array}{c|c|cccc} & X_{n+h} & X_n & X_{n-1} & \cdots & X_1 \\ \hline X_{n+h} & \gamma(0) & \gamma(h) & \gamma(h+1) & \cdots & \gamma(n+h-1) \\ \hline X_n & \gamma(h) & & & & \\ X_{n-1} & \gamma(h+1) & & & & \\ \vdots & \vdots & & \Gamma_n & \to & \\ X_1 & \gamma(n+h-1) & & & & \end{array}$$

③ $E(X_{n+h} - \hat{X}_{n+h})^2 = ?$

$\cancel{\gamma(0)} \rightarrow (\sigma_{YY} - \sigma_{YX}^T \Sigma_{YX}^{-1} \sigma_{YX}) \rightarrow = \gamma(0) - \underset{\sim}{\gamma}_n^T(h) \Gamma_n^{-1} \underset{\sim}{\gamma}_n(h)$

④ $E(X_{n+h} - \hat{X}_{n+h}) = 0$

⑤ $E((X_{n+h} - \hat{X}_{n+h})(X_j) - \mu)) = \text{cov}((X_{n+h} - \hat{X}_{n+h}), (X_j - \mu)) = 0$

$1 \leq j \leq n$

⑥

$$\begin{array}{c|c|c} & Y & \underset{\sim}{X} \\ \hline Y & \sigma_{YY} & \sigma_{Y\underset{\sim}{X}}^T \\ \hline \underset{\sim}{X} & \sigma_{Y\underset{\sim}{X}} & \Sigma_X \end{array}$$

$\text{Cov}((\overset{Y}{X_{n+h}} - \mu), (\overset{X_j}{X_j} - \mu)) = \sigma_{Y \cdot X_j}$

$\text{Cov}((\hat{X}_{n+h} - \mu), (X_j - \mu)) = ?$

$= \text{Cov}(\hat{\underset{\sim}{a}}^T(\underset{\sim}{X} - \underset{\sim}{\mu}), \underset{\sim}{e}_j^T(\underset{\sim}{X} - \underset{\sim}{\mu}))$

$= \underset{\sim}{e}_j^T \underset{\sim}{\gamma}_n(h)$

$= \sigma_{Y, X_j}$

$\because \text{cov}(AX, BX)$

$= A \Sigma_X B^T$

$\underset{\sim}{a}^T = \Gamma_n^{-1} \gamma_n(h)$

$\mathcal{E}(X) = \mathcal{E}(X^T X) \perp$ error

$\downarrow$ orthogonal

Ex. 1 One step prediction: (i.e. $h=1$)

for AR(1) process $X_t = \phi X_{t-1} + W_t$

$W_t \sim WN(0, \sigma_w^2)$.

what is the estimated or predicted value of X_{n+1} when $(X_n, X_{n-1}, \dots, X_1) = \underset{\sim}{X}_n$ given?

$|\phi|<1 \Longrightarrow$ known (say)

$$\Gamma_n \underset{\sim}{a} = \underset{\sim}{\gamma}_n(1) \Leftarrow \hat{\underset{\sim}{a}} = \Gamma_n^{-1} \underset{\sim}{\gamma}_n(1)$$

$$\hat{\underset{\sim}{a}} = \left(\quad \right)$$

$$\begin{array}{c|c|cccc} & X_{n+1} & X_n & \dots & & X_1 \\ \hline X_{n+1} & \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(n) \\ \hline X_n & \gamma(1) & & & & \\ \vdots & \gamma(2) & & & \Gamma_n & \\ \vdots & \vdots & & & & \\ X_1 & \gamma(n) & & & & \end{array}$$

$$\gamma(h) = \frac{\sigma_w^2 \phi^{|h|}}{1-\phi^2}$$

$$\begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \phi^2 & & & & \vdots \\ \vdots & & & & \\ \phi_{n-1} & \dots & & \phi^2\ \phi & 1 \end{pmatrix} \underset{\sim}{a} = \begin{pmatrix} \phi \\ \phi^2 \\ \phi^3 \\ \vdots \\ \phi^n \end{pmatrix} \Rightarrow \therefore \underset{\sim}{a} = \begin{pmatrix} \phi \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where $|\phi|<1$, $\phi \neq 0$

$$E(X_{n+1} - \hat{X}_{n+1})^2 = \gamma(0) - \underset{\sim}{\gamma}_n^T(1)\, \Gamma_n^{-1} \underset{\sim}{\gamma}_n(1)$$

$$= \gamma(0) - \underset{\sim}{\gamma}_n^T(1) \hat{\underset{\sim}{a}}$$

$$= \frac{\sigma_w^2}{1-\phi^2} - (\phi) \frac{\phi \sigma_w^2}{1-\phi^2}$$

$$= \underline{\underline{\sigma_w^2}}$$

(ex. $X_t = 0.79 X_{t-1} + \omega_t$)

Ex. 2. Predict X_{n+h} when, $(X_n, X_{n-1}, \dots, X_1)$ is given for AR(1).

$$\begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & & & & \vdots \\ \phi_{n-1} & \dots & & \phi^2\ \phi & 1 \end{pmatrix} \underset{\sim}{a} = \begin{pmatrix} \phi^h \\ \phi^{h+1} \\ \vdots \\ \phi^{n+h-1} \end{pmatrix}$$

$$\begin{array}{c|c|ccc} & X_{n+h} & X_n & X_{n-1} & \dots \\ \hline X_{n+h} & \gamma(0) & & & \\ \hline X_n & \gamma(h) & & & \\ X_{n-1} & \gamma(h+1) & & \Gamma_n & \\ \vdots & \vdots & & & \end{array}$$

$$\therefore \underset{\sim}{a} = \begin{pmatrix} \phi^h \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$\hat{X}_{n+h} = \phi^h X_n$

$$E(\hat{X}_{n+h} - X_{n+h})^2 = \gamma(0) - \gamma_n^T(h)\, g$$
$$= \frac{\sigma_w^2}{1-\phi^2} - \frac{\sigma_w^2 \cdot \phi^{2h}}{1-\phi^2}$$
$$= \frac{\sigma_w^2(1-\phi^{2h})}{(1-\phi^2)}$$

Trying to predict for long run, $\therefore$ prediction error will incr.
[For AR process].

$x_1, x_2, x_3, \ldots, x_5, x_6 \ldots \rightarrow$ Missing value of data.
Plugg-in.

Missing Notes
12/10/18

(i) Q. To prove, $\exists$ unique "Best Linear Prediction"

Let $\underset{\sim}{\beta}$ and $\underset{\sim}{\theta}$ be both Best Least Square Estimates

$$\left.\begin{array}{l} \Sigma_x \underset{\sim}{\beta} = \sigma_{y.\underset{\sim}{x}} \\ \Sigma_x \underset{\sim}{\theta} = \sigma_{y.\underset{\sim}{x}} \end{array}\right\} \quad \therefore E(\underset{\sim}{\beta}^T \underset{\sim}{x}) = E(\underset{\sim}{\theta}^T \underset{\sim}{x}) \Rightarrow E((\underset{\sim}{\beta}^T - \underset{\sim}{\theta}^T)\underset{\sim}{x}) = 0$$

$$V((\underset{\sim}{\beta}^T - \underset{\sim}{\theta}^T)\underset{\sim}{x}) = (\underset{\sim}{\beta} - \underset{\sim}{\theta})^T \Sigma_x (\underset{\sim}{\beta} - \underset{\sim}{\theta})$$
$$= (\underset{\sim}{\beta} - \underset{\sim}{\theta})^T (\cancel{\Sigma_x \beta - \Sigma_x \theta})^{\underset{\sim}{0}}$$
$$= \underset{\sim}{0}$$

$\therefore (\underset{\sim}{\beta}^T - \underset{\sim}{\theta}^T)\underset{\sim}{x}$ has zero expectation, zero variance.

$\Rightarrow (\underset{\sim}{\beta}^T - \underset{\sim}{\theta}^T)\underset{\sim}{x} = 0$ with probability 1

$\therefore \boxed{\underset{\sim}{\beta}^T = \underset{\sim}{\theta}^T}$ $\quad \therefore$ There is a unique B.L. Prediction

(ii) If $\hat{\underset{\sim}{\beta}}\underset{\sim}{x}$ is the best linear estimator, then.

$$V(Y - \hat{\underset{\sim}{\beta}}^T \underset{\sim}{x}) = \sigma_{Y.\underset{\sim}{x}}$$
$$= V(Y) - V(\hat{\underset{\sim}{\beta}}^T \underset{\sim}{x}) = \sigma_{YY} - \sigma_{Y.\underset{\sim}{x}}^T \Sigma_x^{-1} \sigma_{Y.\underset{\sim}{x}}$$

$$\Sigma = \begin{array}{c} Y \\ \underset{\sim}{X} \end{array} \left[\begin{array}{c|c} \sigma_{YY} & \sigma_{Y.\underset{\sim}{x}}^T \\ \hline \sigma_{Y.x} & \Sigma_x \end{array}\right]$$

$$\Rightarrow |\Sigma| = |\Sigma_x| \, |\sigma_{YY} - \sigma_{Y.\underset{\sim}{x}}^T \Sigma_x^{-1} \sigma_{Y.\underset{\sim}{x}}|$$

$$\therefore \frac{|\Sigma|}{|\Sigma_x|} = \frac{1}{\frac{|\Sigma_x|}{|\Sigma|}} = \frac{1}{(1,1)\text{ Element of } \Sigma^{-1}}$$

$$\because V(Y - \hat{\underset{\sim}{\beta}}^T \underset{\sim}{x}) = \cancel{\sigma_{YY} - \sigma_{Y.\underset{\sim}{x}}^T}$$
$$V(Y) + V(\hat{\underset{\sim}{\beta}}^T \underset{\sim}{x}) - 2\,\mathrm{cov}(Y, \hat{\underset{\sim}{\beta}}^T \underset{\sim}{x})$$
$$= \sigma_{YY} + \hat{\underset{\sim}{\beta}}^T \Sigma_x^{-1} \hat{\underset{\sim}{\beta}} - 2\hat{\underset{\sim}{\beta}}^T \sigma_{Y.\underset{\sim}{x}}$$
$$= \sigma_{YY} + \hat{\underset{\sim}{\beta}}^T \Sigma_x^{-1} \hat{\underset{\sim}{\beta}} - 2(\Sigma_x^{-1} \sigma_{Y.\underset{\sim}{x}})^T \sigma_{Y.\underset{\sim}{x}}$$

$$= \sigma_{YY} + (\Sigma_x^{-1}\sigma_{Yx})\Sigma_x^{-1}(\Sigma_x^{-1}\sigma_{Yx}) - 2(\Sigma_x^{-1}\sigma_{Yx})^T\sigma_{Yx}$$

$$= \sigma_{YY} + \sigma_{Yx}^T\Sigma_x^{-1}\sigma_{Yx} - 2\sigma_{Yx}^T\Sigma_x^{-1}\sigma_{Yx}$$

$$= (\sigma_{YY} - \sigma_{Yx}^T\Sigma_x^{-1}\sigma_{Yx})$$

(iii) $\text{corr}^2(Y, \hat{\beta}^T \underset{\sim}{x}) = \left(\dfrac{\text{cov}(Y, \hat{\beta}^T \underset{\sim}{x})}{\sqrt{V(Y)}\sqrt{V(\hat{\beta}^T \underset{\sim}{x})}}\right)^2$ already known each term

$$= \left(\frac{(\sigma_{Y\underset{\sim}{x}}^T\Sigma_x^{-1}\sigma_{Y\underset{\sim}{x}})^2}{\sigma_{YY}(\sigma_{Y\underset{\sim}{x}}^T\Sigma_x^{-1}\sigma_{Y\underset{\sim}{x}})}\right) = \frac{\sigma_{Y\underset{\sim}{x}}^T\Sigma_x^{-1}\sigma_{Y\underset{\sim}{x}}}{\sigma_{YY}}$$

(iv) $\text{cov}(Y - \hat{\beta}^T\underset{\sim}{x}, x_j) = \text{cov}(Y, x_j) - \text{cov}(\hat{\beta}^T\underset{\sim}{x}, \underset{\sim}{e}_j^T\underset{\sim}{x})$

$$= \sigma_{Yx_j} - \hat{\beta}^T\Sigma_x e_j$$

where $\underset{\sim}{e}_j^T\underset{\sim}{x} = x_j$

$$= \sigma_{Yx_j} - \sigma_{Y\underset{\sim}{x}}^T e_j$$

$\because e_j^T = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ j^{th} term

$$= \sigma_{Yx_j} - \sigma_{Yx_j} = 0$$

Regression v.s.	Time series
• $\underset{\sim}{X}$ was <u>not</u> random	• $\underset{\sim}{X}$ is random
• $\hat{\underset{\sim}{\beta}}$ depended on $Y, \underset{\sim}{x}$	• $\hat{\underset{\sim}{\beta}}$ depends only on covariance terms.

Lets. return to prediction in Time Series

Consider,

$$\underset{\sim}{X} = (X_n, X_{n-1}, \dots X_1)$$

$$Y = X_{n+h}$$

$\{x_t\}$ is stationary, then the linear predictor of X_{n+h}'s:

$$\hat{X}_{n+h} = \mu + \sum_{i=1}^{n} a_i(X_{n+1-i} - \mu),$$

where, $\underset{\sim}{a} = \Gamma_n^{-1}\gamma_n(h)$

• $E(X_{n+h} - \underset{\sim}{a}^T\underset{\sim}{X})^2 = \gamma(0) - \gamma_n^T(h)\Gamma_n^{-1}\gamma_n(h)$

$$\Rightarrow (\sigma_{YY} - \sigma_{Yx}^T\Sigma_x^{-1}\sigma_{Yx})$$

• $E(X_{n+h} - \underset{\sim}{a}^T\underset{\sim}{x}) = 0$

• $\text{Cov}(X_{n+h} - \underset{\sim}{a}^T\underset{\sim}{x}, X_j) = 0$

$$\begin{array}{c|c|c} Y = X_{n+h} & \gamma(0) & \gamma(h)\ \gamma(h+1)\dots \\ \hline X_n & & \\ \vdots & \vdots & \Gamma_n \\ X_1 & & \end{array}$$

ASSIGNMENT

<u>Note</u> Name, Roll no., Subject } 4 pg Report.

24/10/18

→ Let $\{X_t\}$ be a stationary time series with known mean function & ACVF

$$\underbrace{X_1, X_2 \ldots X_{k_1}}_{k_1}, \underline{?}, \underbrace{X_{k_1+2}, X_{k_1+3}, \ldots X_{k_1+k_2+1}}_{k_2}$$

OR

$$\underbrace{X_1, X_2, X_3 \ldots, X_{k_1}}_{\text{known}} \Big| \underbrace{X_{k_1+1}}_{\text{to predict}}, X_{k_1+2} \ldots$$

OR work in reverse dir^n from "k_2" known values.

	Y	$\underset{\sim}{X}$
Y	σ_{YY}	$\sigma_{Y\underset{\sim}{X}}^T$
$\underset{\sim}{X}$	$\sigma_{Y\underset{\sim}{X}}$	Σ_X

$\sim$

	$\underset{\sim}{X_1}$	Y	$\underset{\sim}{X_2}$
$\underset{\sim}{X_1}$	Σ_{X_1}	$\sigma_{Y\underset{\sim}{X_1}}^T$	$\Sigma_{X_1X_2}^T$
Y	$\sigma_{Y\underset{\sim}{X_1}}$	σ_{YY}	$\sigma_{Y\underset{\sim}{X_2}}$
$\underset{\sim}{X_2}$	$\Sigma_{X_1X_2}$	$\sigma_{Y\underset{\sim}{X_2}}^T$	Σ_{X_2}

→ cov.$(\underset{\sim}{X_1}, \underset{\sim}{X_2})$

after dropping rows & column of σ_{YY} position. row parts of σ_{YY}

$$\begin{pmatrix} \sigma_{Y\cdot\underset{\sim}{X_1}} \\ \sigma_{YY} \\ \sigma_{Y\cdot\underset{\sim}{X_2}} \end{pmatrix} = ?$$

$$\underset{\sim}{a} = \Sigma_X^{-1} \underset{\sim}{\sigma}_{Y\cdot X}$$

$$\begin{pmatrix} \sigma_{YY} & \sigma_{Y\cdot\underset{\sim}{X}}^T \\ \sigma_{Y\cdot\underset{\sim}{X}} & \Sigma_X \end{pmatrix}$$

$\sim$

$$\hat{\underset{\sim}{a}} \Rightarrow \begin{pmatrix} \Sigma_{X_1} & \Sigma_{X_1X_2}^T \\ \Sigma_{X_1X_2} & \Sigma_{X_2} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{Y\underset{\sim}{X_1}} \\ \sigma_{Y\cdot\underset{\sim}{X_2}} \end{pmatrix} \Rightarrow \text{Prediction Matrix/vector}$$

∴ Predicted value of X_{k+1} is

$$\hat{X}_{k+1} = \left(\underset{\sim}{a}^T \cdot \begin{pmatrix} X_{k_1} \\ \underset{\sim}{X}_{k_2} \end{pmatrix}\right)$$

OR Instead of considering the whole dataset, take the 3 datapoints preceeding & ~~3 data~~ following the missing point. ∴ $\hat{\underset{\sim}{a}} \Rightarrow \underbrace{\begin{pmatrix} (3\times3) & (3\times3) \\ (3\times3) & (3\times3) \end{pmatrix}^{-1}}_{6\times6} \begin{pmatrix} (3\times1) \\ (3\times1) \end{pmatrix}_{6\times1}$?

Ex. AR(1) $X_t = \phi X_{t-1} + W_t$, $|\phi|<1$, $W_t \sim WN(0, \sigma^2)$.

$X_1 \ \underline{???} \ X_3$

$\left(\dfrac{\sigma^2}{1-\phi^2}\right)$

	X_1	X_2	X_3
X_1	1	ϕ	ϕ^2
X_2	ϕ	1	ϕ
X_3	ϕ^2	ϕ	1

$A \cdot \mathrm{adj} A = |A|$
$\frac{1}{|A|} \mathrm{adj} A = A^{-1}$

$$\therefore \frac{\sigma^2}{1-\phi^2}\begin{pmatrix}1 & \phi^2\\ \phi^2 & 1\end{pmatrix}^{-1}\begin{pmatrix}\phi\\ \phi\end{pmatrix} = \left(\frac{\sigma^2}{1-\phi^2}\right)\underset{\sim}{a}$$

$$\Rightarrow \left(\frac{\sigma^2}{1-\phi^2}\right)\begin{pmatrix}1 & \phi^2\\ \phi^2 & 1\end{pmatrix}\underset{\sim}{a} = \left(\frac{\sigma^2}{1-\phi^2}\right)\begin{pmatrix}\phi\\ \phi\end{pmatrix}$$

$$\Rightarrow \quad \hat{\underset{\sim}{a}} = \frac{\phi^3-\phi}{\phi^4-1}\begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\hat{\underset{\sim}{a}} = \frac{\phi}{1+\phi^2}\begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\therefore \quad \hat{X}_2 = \frac{\phi}{1+\phi^2}\cancel{\begin{pmatrix}1\\ 1\end{pmatrix}} \cdot (X_1 \quad X_3)\begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\hat{X}_2 = \frac{\phi}{1+\phi^2}(X_1+X_3)$$

Prediction error $=$?

$$E(\hat{X}_2 - X_2)^2 = E\left(\left(\frac{\phi}{1+\phi^2}\right)(X_1+X_3) - X_2\right)^2 = ?$$

$X_2 = \phi X_1 + W_2$
$X_3 = \phi(\phi X_1 + W_2) + W_3$
$\quad = \phi^2 X_1 + \phi W_2 + W_3$

$$\left[\begin{aligned} &= \sigma_{YY} - \sigma_{Y\underset{\sim}{X}}^T \Sigma_X^{-1} \sigma_{Y\underset{\sim}{X}} \\ &= \sigma_{YY} - \sigma_{Y\underset{\sim}{X}}^T \cdot \hat{\underset{\sim}{a}} \end{aligned}\right]$$

$$= \left[\frac{\sigma^2}{1-\phi^2}\cdot 1\right] - \left[\frac{\sigma^2 \phi}{1-\phi^2}\right]\begin{pmatrix}1\\ 1\end{pmatrix}^T \cdot \left(\frac{\phi}{1+\phi^2}\right)\begin{pmatrix}1\\ 1\end{pmatrix}$$

$$= \frac{\sigma^2}{1-\phi^2}\left[1 - \frac{2\phi^2}{1+\phi^2}\right]$$

$$= \frac{\sigma^2}{1+\phi^2}$$

HW: $[X_1, \underline{???}, X_3, X_4]$? for AR(1)

** NOTE: **

$$\begin{pmatrix} A & \underset{\sim}{b} \\ \underset{\sim}{c}^T & d \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + \dfrac{A^{-1}\underset{\sim}{b}\,\underset{\sim}{c}^T A^{-1}}{d - \underset{\sim}{c}^T A^{-1}\underset{\sim}{b}} & \dfrac{-A^{-1}\underset{\sim}{b}}{d - \underset{\sim}{c}^T A^{-1}\underset{\sim}{b}} \\ \dfrac{-\underset{\sim}{c}^T A}{d - \underset{\sim}{c}^T A^{-1}\underset{\sim}{b}} & \dfrac{1}{d - \underset{\sim}{c}^T A^{-1}\underset{\sim}{b}} \end{pmatrix}$$

Assume,

$\Gamma_{n-1} \underset{\sim}{a}_{n-1} = \gamma_{n-1}(1)$ has a known solution.

when $\underset{\sim}{a}_{n-1}$ is the coeff. vector of $(X_n \ldots X_2)$.

$$\underset{\sim}{a}_{n-1} = \cancel{(a_{n-1}\ a_{n-2}} \cdots \underbrace{\begin{matrix}(a_1\ a_2 & \ldots & a_{n-1}) \\ (X_n\ X_{n-1} & \cdots & X_2)\end{matrix}}_{n-1} \Bigg| \begin{matrix} a_n ? \\ X_1 \end{matrix}$$

25/10/18

→ Durbin-Levinson Algorithm:

To predict:

(X_{n+1})

by $(X_n, X_{n-1}, \cdots X_2)$ the coefficients $\underset{\sim}{a}_{n-1}$ is known as a sol^n of:

known

$$\Gamma_{n-1}\, \underset{\sim}{a}_{n-1}^{(old)} = \gamma_{n-1}(1).$$

Now we need to include X_1 in the predictor vector.

So we need to estimate:

1. a_n which is the coeff of X_1
2. update $\underset{\sim}{a}_{n-1}$ as $\underset{\sim}{a}_{n-1}^{(new)}$

as a fn of $\underset{\sim}{a}_{n-1}^{(old)}$

for the vector

$(X_n, X_{n-1}, X_{n-2} \cdots X_1)$

Known $\Gamma_{n-1}\, \underset{\sim}{a}_{n-1}^{(old)} = \gamma_{n-1}(1)$

$X_n \to X_2 \quad X_1$

$$\underbrace{\begin{pmatrix} \Gamma_{n-1} & \tilde{\gamma}_{n-1}(1) \\ \tilde{\gamma}^T_{n-1}(1) & \gamma(0) \end{pmatrix}}_{①} \begin{pmatrix} \underset{\sim}{a}_{n-1} \\ a_n \end{pmatrix} = \underbrace{\begin{pmatrix} \underset{\sim}{\gamma}_{n-1}(1) \\ \gamma(n) \end{pmatrix}}_{②}$$

$$\begin{array}{c|c|ccc|c} & X_{n+1} & X_n & \cdots & & X_1 \\ \hline X_{n+1} & \gamma(0) & \gamma(1) & \cdots & & \gamma(n) \\ \hline X_n & \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \vdots & \vdots & \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ & & \vdots & & & \vdots \\ X_1 & \gamma(n) & & & & \gamma(0) \end{array}$$

$$\underset{\sim}{\gamma}_{n-1}(1) = (\gamma(1), \gamma(2) \ldots, \gamma(n-1))$$

$$\tilde{\underset{\sim}{\gamma}}_{n-1}(1) = (\gamma(n-1), \gamma(n-2) \ldots \gamma(1)).$$

$\therefore a_n \Rightarrow$ prdt of last row of $\underset{①}{\underset{\sim}{\Gamma}}$ & $\underset{②}{\text{col. mat.}}$

$$\therefore a_n = \begin{pmatrix} \dfrac{-\underset{\sim}{c}^T A^{-1}}{d - \underset{\sim}{c}^T A^{-1} \underset{\sim}{b}} & \dfrac{1}{d - \underset{\sim}{c}^T A^{-1} \underset{\sim}{b}} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\gamma}_{n-1}(1) \\ \gamma(n) \end{pmatrix}$$

$$a_n = \begin{pmatrix} \dfrac{-\tilde{\gamma}^T_{n-1}(1)\, \Gamma^{-1}_{n-1}}{(\gamma(0) - \tilde{\gamma}^T_{n-1}(1)\, \Gamma^{-1}_{n-1}\, \tilde{\underset{\sim}{\gamma}}_{n-1}(1))} & \dfrac{1}{\gamma(0) - \tilde{\gamma}^T_{n-1}(1)\, \Gamma^{-1}_{n-1}\, \tilde{\underset{\sim}{\gamma}}_{n-1}(1)} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\gamma}_{n-1}(1) \\ \gamma(n) \end{pmatrix}$$

$$a_n = \underline{\underline{\frac{\gamma(n) - \tilde{\underset{\sim}{\gamma}}^T_{n-1}(1)\, \Gamma^{-1}_{n-1}\, \underset{\sim}{\gamma}_{n-1}(1)}{\gamma(0) - \tilde{\underset{\sim}{\gamma}}^T_{n-1}(1)\, \Gamma^{-1}_{n-1}\, \tilde{\underset{\sim}{\gamma}}_{n-1}(1)}}} \quad \text{(known)}.$$

$a_n =$ PACF i.e. $\alpha(n)$

for 'n' lag btw x_{n+1} & x_n

$$a_n = \frac{\gamma(n) - \tilde{\gamma}^T_{n-1}(1)\, \Gamma^{-1}_{n-1}\, \underset{\sim}{\gamma}_{n-1}(1)}{\gamma(0) - \tilde{\gamma}^T_{n-1}(1)\, \Gamma^{-1}_{n-1}\, \underset{\sim}{\gamma}_{n-1}(1)} \longrightarrow \text{soln is known to us.}$$

$\because \Gamma_{n-1}\, \underset{\sim}{a}^{(old)}_{n-1} = \underset{\sim}{\gamma}_{n-1}(1)$

$$a_n = \underline{\frac{\gamma(n) - \tilde{\underset{\sim}{\gamma}}^T_{n-1}(1)\, \underset{\sim}{a}^{(old)}_{n-1}}{\gamma(0) - \underset{\sim}{\gamma}^T_{n-1}(1)\, \underset{\sim}{a}^{(old)}_{n-1}}}$$

[a_n is a PACF of X_1 & X_{n+1} hence $|a_n| \le 1$.]

<u>To show:</u> $\underset{\sim}{a}^{(new)}_{n-1} = \underset{\sim}{a}^{(old)}_{n-1} - a_n\, \Gamma^{-1}_{n-1}\, \tilde{\underset{\sim}{\gamma}}_{n-1}(1)$

$$R \cdot \Gamma_{n-1}\, \underset{\sim}{a}^{(old)}_{n-1} = R \cdot \underset{\sim}{\gamma}_{n-1}(1)$$

$$R\Gamma_{n-1} I\, \underset{\sim}{a}^{(old)}_{n-1} = \tilde{\underset{\sim}{\gamma}}_{n-1}(1)$$

$$(R\Gamma_{n-1}R) R\, \underset{\sim}{a}^{(old)}_{n-1} = \tilde{\underset{\sim}{\gamma}}_{n-1}(1)$$

$$\therefore R\, \underset{\sim}{a}^{(old)}_{n-1} = \Gamma^{-1}_{n-1}\, \tilde{\underset{\sim}{\gamma}}_{n-1}(1)$$

$$\therefore \boxed{\tilde{\underset{\sim}{a}}^{(old)}_{n-1} = \Gamma^{-1}_{n-1}\, \tilde{\underset{\sim}{\gamma}}_{n-1}(1)}$$

Proof:

$$\underset{\sim}{a}_{n-1}^{(new)} = \underset{\sim}{a}_{n-1}^{(old)} - a_n \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)$$

$$= \underset{\sim}{a}_{n-1}^{(old)} - a_n \, \underset{\sim}{\tilde{a}}_{n-1}^{(old)}$$

$$\underset{\sim}{a}_{n-1}^{(new)} = \left(A^{-1} + \frac{A^{-1}\underset{\sim}{b}\,\underset{\sim}{c}^T A^{-1}}{d - \underset{\sim}{c}^T A^{-1}\underset{\sim}{b}} \quad \frac{-A^{-1}\underset{\sim}{b}}{d - \underset{\sim}{c}^T A^{-1}\underset{\sim}{b}} \right) \begin{pmatrix} \underset{\sim}{\gamma}_{n-1}(1) \\ \gamma(n) \end{pmatrix}.$$

$$= \left(\left(\Gamma_{n-1}^{-1} + \frac{\Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1) \underset{\sim}{\tilde{\gamma}}_{n-1}^T(1) \Gamma_{n-1}^{-1}}{\gamma(0) - \underset{\sim}{\tilde{\gamma}}_{n-1}^T(1) \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)} \right) \quad \frac{-\Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)}{\gamma(0) - \underset{\sim}{\tilde{\gamma}}_{n-1}^T(1) \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)} \right) \begin{pmatrix} \underset{\sim}{\gamma}_{n-1}(1) \\ \underset{\text{scalar}}{\gamma(n)} \end{pmatrix}$$

$$= \underbrace{\Gamma_{n-1}^{-1} \underset{\sim}{\gamma}_{n-1}(1)}_{\underset{\sim}{a}_{n-1}^{(old)}} + \frac{\left(\Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1) \underset{\sim}{\tilde{\gamma}}_{n-1}^T(1)\right) \Gamma_{n-1}^{-1} \underset{\sim}{\gamma}_{n-1}(1) - \gamma(n) \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)}{\gamma(0) - \underset{\sim}{\tilde{\gamma}}_{n-1}^T(1) \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)}$$

$$= \underset{\sim}{a}_{n-1}^{(old)} - \frac{\Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1) \left(\gamma(n) - \underset{\sim}{\tilde{\gamma}}_{n-1}^T \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1) \right)}{\gamma(0) - \underset{\sim}{\tilde{\gamma}}_{n-1}^T(1) \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)}$$

$$= \underset{\sim}{a}_{n-1}^{(old)} - \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1) \left(\frac{\gamma(n) - \underset{\sim}{\tilde{\gamma}}_{n-1}^T(1) \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)}{\gamma(0) - \underset{\sim}{\tilde{\gamma}}_{n-1}^T(1) \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)} \right)$$

$$= \underset{\sim}{a}_{n-1}^{(old)} - a_n \Gamma_{n-1}^{-1} \underset{\sim}{\tilde{\gamma}}_{n-1}(1)$$

29/10/18

→ DA Algo: (Prediction interms of Prev. values)

① $\Gamma_{n-1} \underset{\sim}{a}^{(old)}_{n-1} = \gamma^{(1)}_{n-1}$

② $a_n = \dfrac{\gamma(n) - \tilde{\underset{\sim}{\gamma}}^{(1)}_{n-1} \Gamma^{-1}_{n-1} \underset{\sim}{\gamma}^{(1)}_{n-1}}{\gamma(0) - \gamma^{(1)}_{n-1} \Gamma^{-1}_{n-1} \gamma^{(1)}_{n-1}}$

a_n is the PACF between X_1 and X_{n+1} s.t. $\boxed{|a_n| \le 1}$

$\underset{\sim}{a}^{(new)}_{n-1} = \underset{\sim}{a}^{(old)}_{n-1} - a_n \Gamma^{-1}_{n-1} \tilde{\underset{\sim}{\gamma}}^{(1)}_{n-1}$

$\dfrac{E\left(X_{n+1} - \hat{\underset{\sim}{a}}^{T(new)}_{n} \underset{\sim}{X}_n\right)^2}{E\left(X_{n+1} - \underset{\sim}{a}^{T(old)}_{n-1} \underset{\sim}{X}_{n-1}\right)^2} = ?$

$X_{n+1} \leftarrow X_n \cdots X_1$

$X_{n+1} \leftarrow X_n \cdots X_2$

$\sigma_{YY} - \sigma^T_{Y\underset{\sim}{X}} \Sigma^{-1}_X \sigma_{Y\underset{\sim}{X}} \Rightarrow \left(\gamma(0) - \underset{\sim}{\gamma}^{T(1)}_{n-1} \underline{\Gamma^{-1}_{n-1} \underset{\sim}{\gamma}^{(1)}_{n-1}}\right) \rightarrow a^{old(1)}_{n-1}$

$$E\left(X_{n+1} - \underset{\sim}{a}_n^{T(new)} \underset{\sim}{X}_n\right)^2 = \gamma(0) - \underset{\sim}{a}_n^{T(new)} \cdot \underset{\sim}{\gamma}_n(1)$$

$$= \gamma(0) - \begin{pmatrix} \underset{\sim}{a}_{n-1}^{(new)} \\ a_n \end{pmatrix}^T \begin{pmatrix} \underset{\sim}{\gamma}_{n-1}(1) \\ \gamma(n) \end{pmatrix}$$

$$= \gamma(0) - \begin{pmatrix} \underset{\sim}{a}_{n-1}^{(old)} - a_n \Gamma_{n-1}^{-1} \underset{\sim}{\gamma}_{n-1}(1) \\ a_n \end{pmatrix}^T \cdot \begin{pmatrix} \underset{\sim}{\gamma}_{n-1}(1) \\ \gamma(n) \end{pmatrix}$$

$$= \gamma(0) - \left(\underset{\sim}{a}_{n-1}^{T(old)} - a_n \Gamma_{n-1}^{-1} \underset{\sim}{\gamma}_{n-1}^{T}(1)\right)\left(\underset{\sim}{\gamma}_{n-1}(1)\right) - a_n \cdot \gamma(n).$$

$$E\left(X_{n+1} - \underset{\sim}{a}_{n-1}^{T\,old} \cdot \underset{\sim}{X}_{n-1}\right)^2 = \gamma(0) - \underset{\sim}{\gamma}_{n-1}^T(1)\Gamma_{n-1}^{-1}\underset{\sim}{\gamma}_{n-1}(1)$$

$$= \gamma(0) - \underset{\sim}{a}_{n-1}^{T(old)} \underset{\sim}{\gamma}_{n-1}(1) + a_n \tilde{\underset{\sim}{\gamma}}_{n-1}^T(1)\Gamma_{n-1}^{-1}\underset{\sim}{\gamma}_{n-1}(1) - a_n \gamma(n).$$

$$= \left(\gamma(0) - \underset{\sim}{\gamma}_{n-1}^T(1)\Gamma_{n-1}^{-1}\underset{\sim}{\gamma}_{n-1}(1)\right) - a_n\left(\gamma(n) - \tilde{\underset{\sim}{\gamma}}_{n-1}^T(1)\Gamma_{n-1}^{-1}\underset{\sim}{\gamma}_{n-1}(1)\right)$$

$$= \left(\gamma(0) - \underset{\sim}{\gamma}_{n-1}^T(1)\Gamma_{n-1}^{-1}\underset{\sim}{\gamma}_{n-1}(1)\right)\left[1 - a_n^2\right]$$

$$\therefore \boxed{\frac{E\left(X_{n+1} - \underset{\sim}{a}_n^{T(new)} \cdot \underset{\sim}{X}_n\right)^2}{E\left(X_{n+1} - \underset{\sim}{a}_{n-1}^{T(old)} \cdot \underset{\sim}{X}_{n-1}\right)^2} = (1 - a_n^2)} \qquad |a_n| \le 1$$

has to be $\cancel{\ }$ ≥ 0 (non-negative)

If we incr. the number of ~~terms~~ terms, the (expectation) value $\to 0$.
prediction error ratio

$\therefore$ Proportional reduction error with the inclusion of X_1 is $(1-a_n^2) \in (0,1)$.

If we include X_0 for prediction then the ratio will be,

$$(1-a_n^2)(1-a_{n+1}^2) < (1-a_n^2) \in (0,1).$$

X_{n+1} by $(X_n \cdots X_2)$
$(X_n \cdots X_2, X_1)$ $\Big\}$ $(1-a_n^2)$ red^{xn} of η
$(X_n \cdots X_2, X_1, X_0)$. $\Big\}$ $(1-a_n^2)(1-a_{n+1}^2)$ red^{xn} of η

$$\because \frac{E(X_{n+1} - P_{n+1})^2}{E(X_{n+1} - P_{n-1})^2} = \frac{E(X_{n+1} - P_{n+1})^2}{E(X_{n+1} - P_n)^2} \cdot \frac{E(X_{n+1} - P_n)^2}{E(X_{n+1} - P_{n-1})^2}$$

$$= (1 - a_{n+1}^2)(1 - a_n^2).$$

?

$X_n \; X_{n-1} \cdots X_2 \; X_1 \; X_0$
$a_n \; a_{n-1} \cdots a_2 \; a_1 \; a_{n+1} = 0$

→ **Innovation Algorithm:** (Prediction of something in terms of error)

Note: Applicable for all stationary as well as non-stationary time-series

Assumption: ~~Expectation of $(X_t) = 0$~~

~~i.e. $E(X_t) = 0$~~

~~$E(X_t)^2 < \infty$~~

$E(X_t) = 0$

$E(X_t)^2 < \infty$

Notations:

① $k(i,j) = cov(X_i, X_j)$

② $\hat{X}_n = \begin{cases} 0 & , \text{if } n=1 \\ P_{n-1}X_n & , \text{if } n = 2, 3, \dots \end{cases}$

i.e. X_n is predicted as a fn of $(X_{n-1}, X_{n-2}, \dots X_2, X_1)$
Known "n-1" terms

③ (MSE)prediction:

$$\vartheta_n = E(X_{n+1} - P_n X_{n+1})^2$$

epsilon

④ Innovation Terms: one-step prediction error

$$U_n = X_n - \hat{X}_n = X_n - P_{n-1}X_n$$

$$\begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix} = \begin{pmatrix} X_1 - 0 \\ X_2 - \hat{X}_2 \\ X_3 - \hat{X}_3 \\ \vdots \\ X_n - \hat{X}_n \end{pmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{11} & 1 & & \vdots \\ a_{22} & a_{21} & 1 \ddots & \vdots \\ \vdots & & & \\ a_{n-1,n-1} & \cdots & & 1 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{pmatrix}$$

where, $a_{ij} < 0$

$$\underset{\sim}{U}_n = (\underset{\sim}{X}_n - \underset{\sim}{\hat{X}}_n) = A_n \cdot \underset{\sim}{X}_n$$

As $|A_n| = 1$, let $C_n = A_n^{-1}$

$\Rightarrow \underset{\sim}{X}_n = C_n \cdot \underset{\sim}{U}_n$

$$C_n = \begin{pmatrix} 1 & & \cdots & \\ \theta_{11} & 1 & & \vdots \\ \theta_{22} & \theta_{21} & 1 & \vdots \\ \vdots & & \ddots & \\ \theta_{n-1,n-1} & \cdots & & 1 \end{pmatrix}$$

∴ Express $\underset{\sim}{\hat{X}}_n$ in terms of $\underset{\sim}{U}_n$

⇒ ?

Scanned by CamScanner

$\Rightarrow \hat{\underset{\sim}{X}}_n = \underset{\sim}{X}_n - \underset{\sim}{U}_n$

$= C_n \underset{\sim}{U}_n - \underset{\sim}{U}_n$

$= (C_n - I)\underset{\sim}{U}_n$

31/10/18

→ Linear Forecasting : Durbin-Levinson Algo:

Given $X = (X_1, X_2, \dots X_n)^T$ the best linear predictor.

← Slides

→ Innovation Algorithm:

① Applicable to stationary & non-stationary time series

② Predict future values based on estimated one-step errors or innovation.

$$\hat{X}_n = \begin{cases} 0, & \text{if } n=1 \\ P_{n-1} X_n, & \text{if } n \geq 2. \end{cases}$$

$$V_m = E(X_{n+1} - P_n X_{n+1})^2$$

$$U_n = X_n - X_n^2 = \text{innovation}$$

$$= X_n - P_{n-1} X_n$$

- The innovations $(X_{n-j+1} - \hat{X}^{n-j}_{n-j+1})$ are uncorrelated.

$$\hat{X}^n_{n+1} = \begin{cases} 0, & \text{if } n=1 \\ \sum_{j=1}^{n} \theta_{nj} (X_{n-j+1} - \hat{X}^n_{n-j+1}), & \text{o.w.} \end{cases}$$

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} x_1 - 0 \\ x_2 - \hat{x}_2 \\ \vdots \\ x_n - \hat{x}_n \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ a_{11} & 1 & & 0 & \\ a_{22} & a_{21} & 1 & & \\ a_{33} & a_{32} & a_{31} & 1 & \\ \vdots & & & \ddots & \\ a_{n-1,n-1} & \cdots & & & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \end{pmatrix} = A_n \underset{\sim}{X}_n$$

$$\underset{\sim}{U}_n = \underset{\sim}{X}_n - \underset{\sim}{\hat{X}}_n = A_n \underset{\sim}{X}_n$$

$\because$ As $|A_n| = 1$ so A_n is invertible

let $C_n = A_n^{-1}$

$$\underset{\sim}{X}_n = C_n \underset{\sim}{U}_n$$

$$C_n = \begin{pmatrix} 1 & & & & \\ \theta_{11} & 1 & & 0 & \\ \theta_{22} & \theta_{21} & 1 & & \\ \vdots & & & \ddots & \\ \theta_{n-1,n-1} & \cdots & & & 1 \end{pmatrix}$$

$$\underset{\sim}{\hat{X}}_n = \underset{\sim}{X}_n - \underset{\sim}{U}_n$$

$$= C_n \underset{\sim}{U}_n - \underset{\sim}{U}_n$$

$$\boxed{\hat{X}_n = (C_n - I_n)\, \underset{\sim}{U}_n}$$

$$\therefore \hat{X}_n = \Theta_n U_n$$

$$= \Theta_n (X_n - \hat{X}_n)$$

$$\boxed{\hat{X}_n = \begin{pmatrix} 0 & & & & \\ \theta_{11} & 0 & & 0 & \\ \theta_{22} & \theta_{21} & 0 & & \\ \theta_{33} & \theta_{32} & \theta_{31} & 0 & \\ \vdots & & & & \\ \theta_{n-1,n-1} & \cdots & & & 0 \end{pmatrix} (X_n - \hat{X}_n)}$$

using estimated li... of one-step error ~~as~~ $\Rightarrow$ predict the future values.

→ Estimation of Model Parameters. (Yule-Walker).

currently dealing: * ARMA(p,q) w/ ~~...~~ no trend, no seasonality.

- Remove the trend
- Remove the seasonal effect
- Now for weakly stationary process use:

(1) Yule-Walker Estimator: eventually a method of moment estimation process

So equate the theoretical moments

Scanned by CamScanner

← Slides

② MLE:

$\underset{\sim}{X} = (X_1, X_2, \cdots X_n)^T \Rightarrow$ RV's are normally dist^h. WN (as of now.)

Likelihood to be maximized by given ARMA(p,q) parameters.

- Suppose, that $X_1, X_2, \cdots X_n$ is drawn from a zero mean Gaussian ARMA(p,q) process. The likelihood of parameters $\phi \in \mathbb{R}^p$, $\Theta \in \mathbb{R}^q$, and $\sigma_w^2 > 0$ is defined as the density of $\underset{\sim}{X} = (X_1, X_2 \cdots X_n)^T$ under the multivariate Gaussian Model with those parameters.

$$f_X(x_1, x_2 \cdots x_n) = \frac{\exp\left(-\frac{1}{2}(x)^T \Gamma_n^{-1}(x)\right)}{\sqrt{(2\pi)^k |\Gamma_n|}} \rightarrow \frac{\exp\left(-\frac{1}{2}(x)^T \Gamma_n^{-1}(x)\right)}{\sqrt{(2\pi)^k |\Gamma_n|}}$$

$$\begin{cases} v_0 = k(1,1) \\ \Theta_{n,n-k} = v_k^{-1} k(n+1, k+1) - \sum_{j=0}^{k-1} \Theta_{k,k-j} \Theta_{n,n-j} v_j \\ v_n = k(n+1, n+1) - \sum_{j=0}^{n} \Theta_{n,n-j} v_j \end{cases}$$

$$v_0, \Theta_{11}, \Theta_{22}, \Theta_{21}, v_2, \Theta_{33} \cdots$$

Assume mean zero stationary process.

$$\Gamma_n = E(X_n X_n^T)$$

$$L(\Gamma_n) = (2\pi)^{-n/2} |\Gamma_n|^{-1/2} \exp\left\{-\frac{1}{2} \underset{\sim}{X_n}^T \Gamma_n^{-1} \underset{\sim}{X_n}\right\}$$

$$\underset{\sim}{X_n} = C_n \underset{\sim}{U_n} = C_n(\underset{\sim}{X_n} - \hat{\underset{\sim}{X_n}})$$

$X \sim N(\mu, \Sigma)$

$Y = AX \sim N(A\mu, A\Sigma A^T)$

~~$\exp -\frac{1}{2}$~~

$$L(\Gamma_n) = (2\pi)^{-n/2} |\Gamma_n|^{-1/2} \cdot \exp -\frac{1}{2}\left((C_n \underset{\sim}{U_n})^T \Gamma_n^{-1} (C_n \underset{\sim}{U_n})\right) \cdot (J)$$

Jacobian of transformation.

$$\left[\begin{array}{l} \text{corr}(err, x_j) \Rightarrow \text{corr}(Y - \hat{Y}, x_j) = 0 \\ \qquad\qquad Y \leftarrow (x_1, \cdots x_k) \end{array}\right]$$

$\therefore$ $D(U_n) = D_n$ $\qquad \Gamma_n = C_n^T D_n C_n$

$$D_n = \text{diag}(v_0, v_1, v_2, \cdots v_{n-1})$$

Scanned by CamScanner

$$\therefore L(\Gamma_n) = (2\pi)^{-n/2} \left|D_n\right|^{-1/2} \exp\left(\frac{-1}{2}\left(U_n^T C_n^T \Gamma_n^{-1} C_n U_n\right)\right)$$

de

$$\left[f(x) \Rightarrow f(\phi^{-1}(y)) \left\| \frac{\partial y}{\partial x} \right\| \right]$$

$$\Rightarrow (2\pi)^{-n/2} \prod_{i=0}^{n-1} \vartheta_i \exp\left(\frac{-1}{2}\left(\sum_{i=1}^{n} \frac{U_i^2}{\vartheta_{i-1}}\right)\right)$$

$$L(\Gamma_n) \Rightarrow \prod_{i=1}^{n} \frac{e^{-\frac{1}{2}\left(\frac{U_i^2}{\vartheta_{i-1}}\right)}}{\sqrt{2\pi}\,\vartheta_{i-1}}$$

$U_i = X_i - \hat{X}_i$

$\phi_1 \phi_2 \cdots \phi_p \Rightarrow$ AR

$\theta_1 \theta_2 \cdots \theta_q \Rightarrow$ MA

$\sigma_w^2 \rightarrow var(y)$

predicted part has these parameters

Maximize the likelihood
sometimes not possible

$\therefore$

General form:

$$-\ln L(\underset{\sim}{\phi}, \underset{\sim}{\theta}, S(\underset{\sim}{\phi}, \underset{\sim}{\theta}), \sigma^2). \text{ maximized}$$

But where do we stop?

? "AIC": Akaike Information Criteria

"BIC": Bethian Information Criteria

Till here ENDSEMS ↑

1/11/18

→ Integrale AutoRegressive Moving Average Process

ARMA (p,q)   AR(p) degree MA(q)

$$\boxed{\text{ARIMA}(p, d, q)}$$

(stationary process)   Take notes from A...

For $p, d, q \geq 0$, we say a time series $\{X_t\}$ is an ARIMA(p,d,q) process, if $Y_t = (I-B)^d X_t$ is ARIMA(p...

we can write,

$$\boxed{\phi_p(B)\nabla^d X_t = \theta_q(B) W_t}$$

∇ 2nd order diff. $(\underset{\sim}{\theta}_p, \underset{\sim}{\phi}_p)$

$X_1, X_2 > X_2 - X_1 = (I-B)X_2$

$X_2, X_3 > X_3 - X_2 = (I-B)X_3$

$X_3, X_4 > X_4 - X_3 = (I-B)X_4$

$X_4, X_5 > X_5 - X_4 = (I-B)X_5$

$\vdots$

$X_n > X_n - X_{n-1} = (I-B)X_n$

$> \nabla X_3 - \nabla X_2 > \nabla^2 X_3$

$> \nabla X_4 - \nabla X_3 > \nabla^2 X_4$

$> \nabla X_5 - \nabla X_4 > \nabla^2 X_5$

$\vdots$

$> \nabla X_n - \nabla X_{n-1} > \nabla^2 X_n$

$$\begin{aligned}
\nabla X_3 - \nabla X_2 &= (I-B)X_3 - (I-B)X_2 \\
&= X_3 - X_2 - X_2 + X_1 \\
&= X_3 - 2X_2 + X_1 \\
&= (I - 2B + B^2)X_3 \\
&= (I-B)^2 X_3
\end{aligned}$$

• ACF and Autocorrelation Coefficient.

$$X_t = \mu t + \sum_{i=0}^{t} \omega_i$$

$$X_s = \mu s + \sum_{i=0}^{s} \omega_i \qquad \underline{\underline{t \geq s}}$$

$$\text{cov}(X_t, X_s) = \text{cov}\left(\sum_{i=0}^{t} \omega_i, \sum_{i=0}^{s} \omega_i\right) \quad \text{Assume } \omega_0 = 0$$

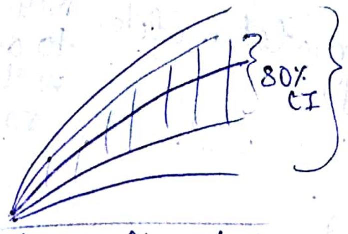
$$= \text{cov}\left(\sum_{i=1}^{t} \omega_i, \sum_{i=1}^{s} \omega_i\right)$$

$$= \text{cov}\left(\sum_{i=1}^{s} \omega_i + \sum_{i=s+1}^{t} \omega_i, \sum_{i=1}^{s} \omega_i\right)$$

$$= \cancel{\text{cov}}\; s\sigma^2 \Rightarrow \boxed{\min\{s,t\} \cdot \sigma^2 = \text{cov}(X_t, X_s)}$$

$$\boxed{\text{corr}(X_s, X_t) = \frac{\min\{s,t\}\sigma^2}{\sqrt{(s \cdot \sigma^2)(t \cdot \sigma^2)}} = \frac{\min\{s,t\}}{\sqrt{st}}}$$

* Brownian motion has same covariance & correlation structure



* Identify & ~~Decide~~ whether data is well-behaving or not

If not, then we need to make according adjustments.

* Look at σ^2 value of R: predict. it may be faulty system bug.

log likelihood
AIC
BIC
σ^2
AICc ⟹ when sample size is less
(corrected)

ex. AIC = 2847.92 ..
AICc = 2847.96

small, ∵ dataset is large. n = 1000.

→ AIC (Akaike Information Criterion).

∵ Most likely data is not well-behaving

∴ Find dist. b/wn Best Predictable Model & Dataset (inref. of having the ~~True Model~~)

we can estimate the ~~true model~~ model, even if we DON'T have the true model.

* Estimates relative quality of statistical models on measu[ring] relative information loss.
* It gives a mean of model selection.
* It is a trade-off btwn <u>goodness of fit</u> & <u>simplicity</u> of t[he] model.

$$\boxed{AIC = 2k - 2\log_e(\hat{L})}$$ ⟹ Minimum is <u>BEST</u>

∵ It computes the "loss of information" ⇓ Measurement of "Entropy", "Randomness", "Uncertainty" of the system ⇓ Nature tries to always withhold this randomness.

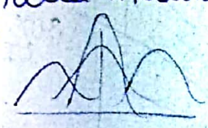
k = number of estimated parameters

$\hat{L}$ = Maximum of likelihood.

gives the flexibility of changing the estimating parameters

* <u>Let</u>, f is the true model.

g_1 & g_2 are the other estimated models (pdf)



* <u>Kullback-Leibler divergence</u>:

$$\boxed{D_{KL}(f\|g_1) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g_1(x)}\, dx} \geq 0$$

~ dist.
∴ will not consider the -ve part

~ <u>NOT</u> symmetric fn.
∴ divergence is emphasized.

$\boxed{\underline{\text{P.T:}}\ D_{KL}(f\|g_1) \geq 0}$ <u>HW</u>

$$= -\int f(x)\log(g_1(x))dx + \int f(x)\log(f(x))dx$$

$$= -\underbrace{\left(-\int f(x)\log f(x)\, dx\right)}_{} + \underbrace{\left(-\int f(x)\log g_1(x)\, dx\right)}_{}$$

⟹ "Shannon Metric of Entropy"

$$= -\underset{f}{H}(f) + \underset{f}{H}(g_1)$$

↓ Relative Entropy of "g_1" in terms of "f"

** CLT: physical significance, any pdf (sample size $n \to \infty$) converges to normal distbn.

pdf 1: $N(\mu, \sigma^2)$

pdf 2: $X(\mu, \sigma^2)$

$\downarrow$

any Non-normal distbn.

The normal pdf represents, converges as the upper bound, and is the most random, or ~~entropy~~ pdf with most entropy.