$\{X_t\}_{-\infty}^{\infty}$ is a weakly stationary time series and $E[X_t]=0 \ \forall t$

<u>Note:</u> $P_n(X_{n+h})$ is projection of X_{n+h} to the ~~span~~ space span $\{X_n, \dots, X_1\}$

So $P_n(X_{n+h})$ is best linear projector

<u>iff</u> it is projector of X_{n+h} to the span

<u>iff</u> $P_n(X_{n+h})$ is a linear func of $\{X_n, \dots, X_1\}$

and $E[(X_{n+h} - P_n(X_{n+h}))X_{n-j}] = 0 \quad \forall j = 0, 1, 2, \dots, n-1$

Note $E[X_t] = 0 \ \forall t$

Consider L^2 space.

Now consider $\bar{P}_n(X_{n+h})$ best linear predictor of X_{n+h} on the basis of $\{X_n, X_{n-1}, \dots, X_1, X_0, X_{-1}, X_{-2}, \dots\}$

Best linear projector may not be of the form $\sum_{j=0}^{\infty} \ell_j X_{n-j}$.

But it should belong to $Sp\{X_n, X_{n-1}, \dots, X_1, X_0, X_{-1}, \dots\} = K_n$

$\bar{P}_n(X_{n+h}) = \underset{u \in K_n}{\arg\min} \, E[(X_{n+h} - u)^2]$ = projection of X_{n+h} onto K_n

<u>Result:</u> From a vector of (X_{n+h}) to a closed subspace (K_n) of a L^2 space always exists

~~So~~ $\bar{P}_n(X_{n+h})$ is best linear predictor

<u>iff</u> $\bar{P}_n(X_{n+h})$ is projector of X_{n+h} to K_n

<u>iff</u> $\bar{P}_n(X_{n+h}) \in K_n$ and $E[(X_{n+h} - \bar{P}_n(X_{n+h}))X_{n-j}] = 0 \quad \forall j \geq 0$

We shall use this to prove the following <u>decomposition theorem</u>

<u>Thm</u>. If $\{X_n\}_{-\infty}^{\infty}$ is a weakly stationary time series, then $\exists$ $WN(0,\sigma^2)$, $\{Z_n\}_{-\infty}^{\infty}$ s.t.

$$X_n = \sum_{j=-\infty}^{\infty} \psi_j Z_{n-j} + V_n \quad \text{when } V_n \in \bigcap_{j\geq 1} K_{n-j} \text{ and}$$

$$\bar{P}_{n-1} V_n = \bar{P}_{n-2} V_n = \cdots = V_n$$

and $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$

- Here V_n is called the deterministic component and $\sum_{j=-\infty}^{\infty} \psi_j Z_{n-j}$ is non-deterministic component.

$$\boxed{\begin{array}{l} X_n = Y_n + U, \quad Y_n \overset{iid}{\sim} N(0,1),\ U \sim N(0,1) \text{ indep.} \\ U \in K_{n-j} \ \forall j \geq 1 \\ [\lim_{m\to\infty} \frac{1}{m}\sum_{\ell=0}^{m-1} X_{n-\ell-j} = U] \text{ a.e.} \end{array}}$$

$U \to$ deterministic component

$Y_n \to$ non-deterministic component

- V_n is a function of n only.

$\longrightarrow$

Consider X_{n+1}, $\bar{P}_n(X_{n+1})$, $\bar{P}_{n-1}(X_{n+1})$, ...

Define,

$$Y_{n,1} = X_{n+1} - \bar{P}_n(X_{n+1})$$

$$Y_{n,2} = \bar{P}_n(X_{n+1}) - \bar{P}_{n-1}(X_{n+1})$$

$$Y_{n,3} = \bar{P}_{n-1}(X_{n+1}) - \bar{P}_{n-2}(X_{n+1})$$

and so on.

From the property of projection

$Y_{n,1}, Y_{n,2}, \ldots$ are orthogonal.

and $X_{n+1} = (Y_{n,1} + Y_{n,2} + \cdots) + V_n$ where

$V_n \in \bigcap_{j\geq 1} K_{n-j}$ i.e. $\bar{P}_{n-j} V_n = V_n \ \forall j \geq 1$

Claim: $y_{n,2}$ and $y_{n-1,1}$ are multiple of same vector L where

$K_n = L \oplus K_{n-1}$

Since K_n and K_{n-1} has dimension difference 1 at most

and $y_{n,2}$ and $y_{n-1,1}$ are both orthogonal to K_{n-1}

$\therefore y_{n,2}$ and $y_{n-1,1}$ are multiple of same vector L where

$K_n = L \oplus K_{n-1}$

$\therefore x_{n+1} = \sum_{j=-\infty}^{\infty} \psi_j Z_{n-j} + V_n$ where Z_t are obtained from L (unique)

$\therefore$ By scaling we get $Var(Z_t) = \sigma^2$ and we get ψ_j

Note: ψ_j depends only on correlation structure of $(X_{n+1}, X_n, X_{n-1}, \dots)$

which is same if we consider $(X_{n-k}, X_{n-k-1}, \dots)$ ①

So ψ_j doesn't depend on (k).

D.

Thm. If $\{X_t\}_{-\infty}^{\infty}$ is weakly stationary then for some $\{Z_t\}_{-\infty}^{\infty}$ we have

$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ ~~where~~ $+ V_t$, where $V_t \in \bigcap_{j=1}^{\infty} K_{t-j}$ i.e.

$\bar{P}_{n-1} V_n = \bar{P}_{n-2} V_n = \dots = V_n$

We shall use the thm to show

Result: If $\{X_t\}$ is stationary and q-correlated (for $q \geq 1$)

then $\{X_t\}_{-\infty}^{\infty}$ is a MA(q)

Hint: Getting decomposition of X_t according to the Thm.

We get $\{Z_t\}_{-\infty}^{\infty}$ and V_t

$Y_{n,j} \in K_{n-j+2}$ (depending on $X_{n-j+2}, X_{n-j+1}, \ldots\ldots$)

$\therefore Y_{n,j} \perp X_{n+1}$ for $j > q+2$ (check!)

$\therefore \psi_j = 0 \quad \forall \; j > q+2$

Now need to show $V_t \equiv 0$

$\bar{P}_{t-q-j} V_t = V_t$ [linear func $\in K_{t-j-q}$]

$Cov(X_t, V_t) = 0$

$\Rightarrow V_t \equiv 0$

Thm: (Wald's Decomposition Thm):

If $\{X_t\}_{-\infty}^{\infty}$ is a stationary (weakly) time series, we can get $\{Z_t\}_{-\infty}^{\infty}$ $WN(0,\sigma^2)$ s.t.

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t \quad \text{where} \quad \sum_{0}^{\infty} \psi_j^2 < \infty \quad \text{and} \quad \psi_j \in \mathbb{R} \ \forall j$$

(Assume, $E(X_t)=0$)

and V_t is deterministic component s.t.

$\bar{P}_{t-1} V_t = \bar{P}_{t-2} V_t = \cdots = V_t$ and

$V_t \in \bigcap_{j=1}^{\infty} K_{t-j}$, $K_{t-j} = \overline{sp\{X_{t-j}, X_{t-j-1}, \ldots\}}$

[Our proof in last classes goes through to prove Wald's Decomposition Thm.]

Corollary: If $\{X_t\}_{-\infty}^{\infty}$ is q-correlated then it is a MA(q) process.

$$\rightarrow \quad X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$$

From the proof of Wald's Decomposition Thm,

$$Z_{t-j} \in K_{t-j} = \overline{sp\{X_{t-j}, X_{t-j-1}, \ldots\}}$$

$$Cov(X_t, Z_{t-j-q}) = 0 \quad \forall j \geq 1$$

$$, \Rightarrow \psi_{q+j} = 0 \quad \forall j \geq 1$$

Only thing to show, $V_t \equiv 0$, $\forall t$

$$V_t \in \bigcap_{j=1}^{\infty} K_{t-j}$$

$$\Rightarrow V_t \perp X_s, \ \forall s \leq t$$

$$\Rightarrow V_t \perp V_t \quad \Rightarrow Var(V_t) = 0 \quad \Rightarrow V_t \equiv 0$$

Eg: $X_t = U + Y_t \quad , t \in \mathbb{Z}$

$Y_t \overset{iid}{\sim} N(0,1)$

U indep. to $\{Y_t\}$

$\rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i = U + \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{L^2} U$

$\therefore \; U \in K_{t-j} \; , \forall j \geq 1$

$\Rightarrow U \in \bigcap_{j \geq 1} K_{t-j}$

$Cov((X_t - U), X_{t-j}) = Cov(Y_t, X_{t-j}) = 0$

$\Rightarrow \bar{P}_{t-1}(X_t) = U$

Note:

$\psi_j = \begin{cases} 1 & \text{for } j=0 \\ 0 & \text{ow} \end{cases}$

and $Z_t \equiv Y_t$

Eg

$X_t = A\cos(\omega t) + B\sin(\omega t) \; , \; \omega \in \mathbb{R}$ const.

$\rightarrow \quad Cov(X_t, X_{t+h}) = \gamma(h) = \cos(\omega h)$

Deterministic component of X_t is X_t itself.

$E(A) = E(B) = 0$

$Var(A) = Var(B) = 1$

$Cov(A,B) = 0$

$2(\cos(\omega))\, X_{t-1} - X_{t-2}$

$= 2(\cos(\omega))\, \{A\cos(\omega(t-1)) + 2B(\cos(\omega))\sin(\omega(t-1)) - X_{t-2}$

$= A\cos(\omega t) + B\sin(\omega t) + A\cos(\omega(t-2)) + B\sin(\omega(t-2)) - X_{t-2}$

$= X_t$

$\therefore \bar{P}_{t-1}(X_t) = \bar{P}_{t-1}(2\cos(\omega) X_{t-1} - X_{t-2})$

$= 2\cos(\omega) X_{t-1} - X_{t-2} = X_t$

Simillarly,

$\bar{P}_{t-j} X_t = X_t \;, j \geq 1$

$X_t \in \bigcap_{j=1}^{\infty} K_{t-j}$

$\Rightarrow X_t$ is deterministic.

Def^n: $p \geq 1$, Auto Regressive process of order p, $AR(p)$ is a stationary solution to

$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$, where $\phi_i \in \mathbb{R}$ const.

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

and $Cov(X_s, Z_t) = 0 \;\forall s < t$

- $P_t(X_{t+1})$, $t \geq p$

$= P_t(\phi_1 X_t + \cdots + \phi_p X_{t-p+1} + Z_{t+1})$

$= \phi_1 P_t(X_t) + \phi_2 P_t(X_{t-1}) + \cdots + \phi_p P_t(X_{t-p+1}) + 0$ [Since $Cov(Z_t, X_l) = 0 \;\forall t > l$]

$= \phi_1 X_t + \cdots + \phi_p X_{t-p+1}$

We could have get it by

$$Cov\left((X_{t+1} - P_t(X_{t+1})),\ X_{t-j} \right),\quad j \geq 0$$

$$= Cov\left(Z_{t+1},\ X_{t-j} \right) = 0,\quad \forall j \geq 0$$

Eg 4. $P_t(X_{t+h})$, $h \geq 1$ in AR(1) process

$E(X_t) = 0$

$$P_t(X_{t+h}) = P_t(\phi X_{t+h-1} + Z_t) = \phi P_t(X_{t+h-1}) = \phi^h X_t \text{ (by induction)}$$

We could have obtain it by sol^n to eq^n

$$\Gamma_n \underset{\sim}{a} = \gamma_n(h)$$

$$\Gamma_n = \begin{pmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \phi^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \cdots & 1 \end{pmatrix} \qquad \gamma_n(h) = \begin{pmatrix} \phi^h \\ \phi^{h+1} \\ \vdots \\ \phi^{h+n} \end{pmatrix}$$

$$\underset{\sim}{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

$$\underset{\sim}{a}' = (\phi^h, 0, 0, \ldots, 0)$$

Let $\{Y_t\}$ be AR(1)

$E(Y_t) = \mu \neq 0$

$X_t = Y_t - \mu$ is AR(1) with mean 0

For $h \geq 1$,

$$P_t(X_{t+h}) = \phi^h X_t$$

$$\Rightarrow P_t(Y_{t+h} - \mu) = \phi^h (Y_t - \mu)$$

$$\Rightarrow P_t(Y_{t+h}) = \phi^h Y_t + (1 - \phi^h)\mu$$

$$= \mu + \phi^h (Y_t - \mu)$$

2.11. In a sample of size 100 from an AR(1) process with mean $=\mu$, $\phi=0.6$ $\sigma^2=2$. From the data $x_1, x_2, \dots, x_{100}$; we got $\bar{x}_{100} = 0.271$. Construct approximate 95% confidence interval for μ. Hence test $H_0: \mu=0$

$$X_t = \phi X_{t-1} + Z_t \text{ , where } Z_t \sim WN(0,\sigma^2)$$

$$E[Y_t]=0$$

$$X_t = Y_t + \mu$$

2.12. $x_1, \dots, x_{100}$ data from MA(1) with mean $=\mu$. $\theta=-0.6$, $\sigma^2=1$ and $\bar{x}_n = 0.157$ Consider approximate 95% confidence interval for μ. Hence test $H_0: \mu=0$

$$Y_t = \theta Y_{t-1} + Z_t \quad \text{where } Z_t \sim WN(0,\sigma^2)$$

$$E[Y_t]=0$$

$$X_t = Y_t + \mu$$

We need sampling distn of $\bar{x}_{100}$ ($\hat{\mu}$?)

We shall do it and also we shall get estimate $\hat{\gamma}(h)$ of auto-covariance function and get their sampling distn.

$\{X_t\}$ stationary time series with mean μ and ACVF $\gamma(h)$

We estimate $\hat{\mu} = \bar{x}_n = \frac{1}{n}(x_1 + \dots + x_n)$

Error $= E[(\bar{x}_n - \mu)^2] \to 0$ iff $\bar{x}_n$ is consistent estimator of μ.

$$E[(\bar{x}_n - \mu)^2] = \frac{1}{n^2} E\Big[\sum_{i=1}^{n} (x_i - \mu)\Big]^2$$

$$= \frac{1}{n^2} \sum_i \sum_j \gamma(i-j)$$

$$= \frac{1}{n^2} \sum_{|h|<n} (n-|h|)\, \gamma(h)$$

Result: If $\gamma(h) \to 0$ as $h \to \infty$, then $E(\text{error}^2) \to 0$ as $n \to \infty$

$\rightarrow \quad 0 \leq E(\text{error}_n^2) \leq \frac{1}{n} \sum_{|h|<n} |\gamma(h)| \leq \frac{2}{n} \sum_{h=0}^{n} |\gamma(h)| \longrightarrow 0$

as $|\gamma(h)| \rightarrow 0$

Result 2: $n\, E[(\bar{x}_n - \mu)^2] \longrightarrow \sum_{|h|<\infty} \gamma(h)$ whenever $\sum_{|h|<\infty} |\gamma(h)| < \infty$

$\rightarrow \quad n\, E[(\bar{x}_n - \mu)^2] = \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma(h) \rightarrow \sum_{|h|<\infty} \gamma(h)$

[Abel's convergence Thm.]

If $\{X_t\}$ is AR(p) or MA(q) process with $\{Z_t\} \overset{iid}{\sim} N(0, \sigma^2)$,

OR If $\{X_t\}$ is Gaussian process,

$\Rightarrow \quad \sqrt{n}\, (\bar{X}_n - \mu) \sim AN(0, v)$ where $v = \sum_{|h|<\infty} \gamma(h)$

95% CI for μ is given by $\left(\bar{x}_n - \frac{1.96}{\sqrt{n}} \sqrt{v},\ \bar{x}_n + \frac{1.96}{\sqrt{n}} \sqrt{v}\right)$

Eg. 1 $\quad Y_t \sim AR(1)$, $X_t = Y_t + \mu$

$E(X_t) = E[Y_t] + \mu = \mu$

$$v = \sum_{|h|<\infty} \gamma(h) = 2\left(\sum_{h=1}^{\infty} \phi^h\right) \frac{\sigma^2}{1-\phi^2} + \frac{\sigma^2}{1-\phi^2}$$

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

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

Eg 2. $\{Y_t\} \sim MA(1)$

$$X_t = Y_t + \mu$$

$$E(X_t) = \mu + 0 = \mu$$

$$v = \sum_{|h|<\infty} \gamma(h) = \gamma(0) + 2\gamma(1)$$

$$= (1+\theta^2)\sigma^2 + 2\theta\sigma^2$$

$$= (1+\theta)^2\sigma^2$$

$$\sqrt{v} = |1+\theta|\sigma$$

Sample ACVF ($\widehat{\gamma(h)}$)

$\{X_t\}$ stationary time series

Data : $x_1, x_2, \dots, x_n$ is given

$$\widehat{\gamma(h)} = \frac{1}{n} \sum_{i=1}^{n-|h|} (x_i - \bar{x}_n)(x_{i+|h|} - \bar{x}_n)$$

Note: Here in the estimate, we used divisor n and considered mean $\bar{x}_n$ [But if we did not do it, the ACVF matrix $\hat{\Gamma}_{n\times n}$ may not be nnd]

Thm: $\hat{\Gamma}_{n\times n}$ thus obtained is nnd.

Corollary: $\widehat{\rho(h)} = \dfrac{\widehat{\gamma(h)}}{\widehat{\gamma(0)}}$, then $((\widehat{\rho(i-j)}))_{i,j=1}^{n}$ is n.nd

→ Define $y_i = x_i - \bar{x}_n$, $i = 1(1)n$

$= 0$, $i > n$

Take $k > n$, $(y_1, \dots, y_k)$

Define,

$$T_{k\times 2k} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 & y_1 & y_2 & \cdots & y_k \\ 0 & 0 & 0 & 0 & \cdots & 0 & y_1 & y_2 & y_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_1 & y_2 & y_3 & \cdots & y_{k-2} & y_{k-1} & y_k & 0 & \cdots & & 0 \end{bmatrix}$$

Note that

$$\hat{\Gamma}_{k\times k} = \frac{1}{n} TT'$$

Hence $\hat{\Gamma}_{k\times k}$ is nnd

$\hat{\Gamma}_{n\times n}$ is a principle submatrix of $\hat{\Gamma}_{k\times k}$.

$\therefore$ $\hat{\Gamma}_{n\times n}$ is nnd.

<u>Result</u>: If some $\gamma_i \neq 0$, then $\hat{\Gamma}_{k\times k}$ is p.d.

→ If $\gamma_1 \neq 0$, The submatrix of T, with column's 2nd, 3rd, ... (k+1)th, is nonsingular

$\therefore$ Rank(T) = k

$\Rightarrow$ $\hat{\Gamma}_{k\times k}$ is non-singular i.e. p.d.

If $\gamma_1 = 0$, $\gamma_2 \neq 0$, the submatrix of T with columns 3rd, 4th, ..., (k+2)th, is non-singular.

$$\text{Rank}\left(\hat{\Gamma}_{k\times k}\right) = k$$

$\Rightarrow$ $\hat{\Gamma}_{k\times k}$ is pd

Similarly we have $\hat{\Gamma}_{k\times k}$ is pd if any $\gamma_j \neq 0$

$\therefore$ $\hat{\Gamma}_{n\times n}$ is also pd.

- Time Series data $\rightarrow x_1, x_2, \ldots, x_{100}$

We obtain $\widehat{\rho(1)} = 0.438$, $\widehat{\rho(2)} = 0.145$ [Assume iid noise]

(a) Assuming the data generated from AR(1) model, construct the approximate 95% confidence interval for both $\rho(1)$ and $\rho(2)$. Based on these two CI, are the data consistent with AR(1) with $\phi = 0.8$.

(b) Assuming the data generated from MA(1) model, construct 95% confidence interval for both $\rho(1)$ and $\rho(2)$ based on these two CI. Are the data consistent with MA(1) model with $\theta = 0.6$.

We need assymptotic variance of $\widehat{\rho(1)}$ and $\widehat{\rho(2)}$ under the model. Estimate $\hat{\phi}$ from $\widehat{\rho(1)}$ and $\widehat{\rho(2)}$, use it in the expression for assymp. variance.

Time series data $x_1, x_2, \ldots, x_n$

Sample ACVF $\widehat{\gamma(h)}$, $h = 1, 2, \ldots, k$ with $k \ll n$

$\boxed{\text{Convention: } n \geq 50,\ k \leq n/4}$

Sample ACF $\widehat{\rho(h)} = \dfrac{\widehat{\gamma(h)}}{\widehat{\gamma(0)}}$, $h = 1, 2, \ldots, k$

Let, $\underset{\sim}{\rho_k} = (\rho(1), \ldots, \rho(k))$

$\underset{\sim}{\hat{\rho}_k} = (\widehat{\rho(1)}, \ldots, \widehat{\rho(k)})$

<u>Bartlett's Formula:</u> $\hat{\rho}_k$ is assymptotically MVN $(\rho_k, \frac{1}{n} W_{k \times k})$, where

$$W_{k \times k} = ((w_{ij}))_{i,j=1}^{k} \quad \text{and} \quad w_{ij} = \sum_{l=1}^{\infty} (\rho(l+i) + \rho(l-i) - 2\rho(l)\rho(i))(\rho(l+j) + \rho(l-j) - 2\rho(l)\rho(j))$$

Works for Linear Process generated by iid noise.

Application 1:

MA(1) Process

$$\omega_{11} = 1 - 3\rho(1)^2 + 4\rho(1)^4 \quad \text{for } i=1$$

$$= 1 + 2\rho(1)^2 \quad \text{for } i \geq 2$$

Application 2:

AR(1) Process

$$\rho(h) = \phi^{|h|}$$

$$\omega_{ii} = \sum_{k=1}^{i} \phi^{2i} (\phi^{-k} - \phi^{k})^2 + \sum_{k=i+1}^{\infty} \phi^{2k} (\phi^{-2i} - \phi^{2i})^2$$

$$= (1 - \phi^{2i}) \frac{(1+\phi^2)}{(1-\phi^2)} - 2i\phi^{2i} \quad \text{(derive)}$$

Application 3:

Testing for iid noise

$$\hat{\rho}_k = (\widehat{\rho(1)}, \ldots, \widehat{\rho(k)}) \sim N_k(0, \tfrac{1}{n} I_k) \quad \text{under iid noise}$$

$$\text{i.e. } \omega_{ij} = 1_{[i=j]}$$

- The result gives us a scope to ~~test~~ iid noise time series

H_0: data is from iid noise

H_1: data is not from iid noise

Method 1:

$$\widehat{\rho(i)} \in \pm \frac{1.96}{\sqrt{n}}, \quad i = 1(1)k$$

95% CI

If more than 5% of i gives $\widehat{\rho(i)} \notin CI$, we reject H_0

Method 2:

$$Q^2 = n \sum_{i=1}^{k} \widehat{\rho(i)}^2 \sim \chi^2_k \quad \text{under } H_0$$

If $Q^2 > \chi^2_{k,0.95}$, we reject H_0 (**Portmanteau test**)

Alternative test by Ljung and Box

Modify the χ^2_k statistics as

$$Q_{LB} = n(n+2) \sum_{i=1}^{k} \widehat{\rho(i)}^2 / n-i \sim \chi^2_k \quad \text{under } H_0$$

This test method based on this statistic is more efficient.