Chapter 2 - Stationary Process

2.1 Basic Properties

Best Predictor: Minimize mean-squared error by choosing $m(X_n)$

Choose m to minimize

$$E[(X_{n+h}-m(X_n))^2] = E_{X_n}[E_{n+h|X_n}[(X_{n+h}-m(X_n))^2 | X_n]]$$

Consider $\min_c E[(Y-c)^2] \Rightarrow c^* = E(Y)$

Thus, $m(X_n)$ that minimize $E_{n+h|X_n}[(X_{n+h}-m(X_n))^2|X_n]$ is

$$m(X_n) = E_{X_{n+h}|X_n}[X_{n+h}|X_n]$$

Assume $\hat{m}(X_n)$ minimizes $E[(X_{n+h}-m(X_n))^2]$.

$$E[(X_{n+h}-\hat{m}(X_n))^2]$$
$$= E[(X_{n+h}-E(X_{n+h}|X_n)+E(X_{n+h}|X_n)-\hat{m}(X_n))^2]$$
$$= E[(X_{n+h}-E(X_{n+h}|X_n))^2] + 2E[(X_{n+h}-E(X_{n+h}|X_n))(E(X_{n+h}|X_n)-\hat{m}(X_n))]$$
$$+ E[(E(X_{n+h}|X_n-\hat{m}(X_n))^2]$$

$$2E_{X_n}E_{X_{n+h}|X_n}[(X_{n+h}-E(X_{n+h}|X_n))(E(X_{n+h}|X_n)-\hat{m}(X_n))|X_n]$$
$$= 2E_{X_n}[(E(X_{n+h}|X_n)-\hat{m}(X_n)$$

$$\Rightarrow E[(X_{n+h}-\hat{m}(X_n))^2] = E[(X_{n+h}-E(X_{n+h}|X_n))^2] + E_{X_n}[(E(X_{n+h}|X_n)-\hat{m}(X_n))^2]$$

>0 unless $E(X_{n+h}|X_n) = \hat{m}(X_n)$

But this obviously extends to conditioning on $(X_1 \dots X_n)$

Choose a model for $E(X_{n+h}|X_n)$, always start with <u>linear</u>.

$$E[(X_{n+h}-(a+bX_n))^2] = E(X_{n+h}^2) - 2E[X_{n+h}(a+bX_n)] + E[(a+bX_n)^2]$$
$$= E(X_{n+h}^2) - 2(aE(X_{n+h}) + bE(X_{n+h}X_n)) + a^2 + b^2E(X_n^2) + 2abE(X_n)$$

$$\begin{cases} \frac{\partial}{\partial a}: -2E(X_{n+h}) + 2a + 2bE(X_n) = 0 \\ \frac{\partial}{\partial b}: -2E(X_{n+h}X_n) + 2aE(X_n) + 2bE(X_n^2) = 0 \end{cases} \Rightarrow \begin{cases} \hat{a} = \mu_{n+h} - \hat{b}\mu_n \\ \hat{b} = \frac{Cov(X_{n+h}, X_n)}{Var(X_n)} \end{cases}$$

μ_{n+h} μ_n

$$\hat{X}_{n+h} = \mu_{n+h} + \frac{Cov(X_{n+h}, X_n)}{Var(X_n)}(X_n - \mu_n)$$

If $\{X_t\}$ is stationary, then $\hat{X}_{n+h} = \mu + \frac{\gamma(h)}{\gamma(0)}(X_n - \mu_n) = \rho(h)X_n + (1-\rho(h))\mu_n$

Properties of $\gamma(h)$

① $\gamma(0) \geq 0$ Variance

② $|\gamma(h)| \leq \gamma(0)$ Cauchy-Schwarz Inequality $|\langle u,v\rangle|^2 \leq \langle v,u\rangle \cdot \langle v,u\rangle$
Define $\langle u,v\rangle = E(uv)$ for $E(u)=E(v)=0$

③ $\gamma(h)$ is even $\gamma(h)=\gamma(-h)$

④ $\gamma(h)$ is non-negative definite

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma(i-j) a_j \geq 0 \quad \forall n \in \mathbb{Z}^+ \text{ and } a \in \mathbb{R}^n.$$

Even stronger property Let $\gamma(h)$ be defined on $h \in \mathbb{Z}$

$\gamma(h)$ non-negative definite and even $\Leftrightarrow$ It is the auto-covariance function of some stationary sequence.

- To verify that a given function is nonegative definite, it is often simpler to find a stationary process that has the given function as its ACVF than to verify the conditions directly.

e.g. $\gamma(h) = \cos(wh)$ is nonnegative definite, since it is the ACVF of the stationary process $X_t = A\cos(wt) + B\sin(wt)$ where A, B are uncorrelated random variables with mean 0 and variance 1.

Strictly Stationary Time Series

$$(X_1 \dots X_n) \overset{d}{=} (X_{1+h} \dots X_{n+h}) \quad \forall\, h \text{ and } n.$$

Two random vectors have the same joint distribution function.

Properties

① All elements of $\{X_t\}$ are identically distributed.

② $(X_t, X_{t+h}) \overset{d}{=} (X_1, X_{1+h})$

③ If $E(X_t^2)<\infty$, then $\{X_t\}$ is weakly stationary

④ Weakly stationary doesn't imply strictly stationary

⑤ iid process is strictly stationary

How to make a stationary sequence ?

Let $\{Z_t\}$ be an iid sequence of random variables.

Let $X_t = g(Z_t, Z_{t-1} \cdots Z_{t-q})$

$\Rightarrow X_t$ is strictly stationary because $(Z_{t+h} \cdots Z_{t+h-q}) \overset{d}{=} (Z_t \cdots Z_{t-q})$

Time series $\{X_t\}$ is q-dependent.

i.e. X_t and X_s are independent if $|t-s|>q$

Note: An iid sequence is 0-dependent.

Generalize to weakly stationary

Adopting a second-order viewpoint, say that $\{X_t\}$ is q-correlated if

$$Cov(X_t, X_s) = 0 \quad \forall \ |t-s|>q$$

$$\text{or} \quad \gamma(h) = 0 \quad \forall \ |h|>q$$

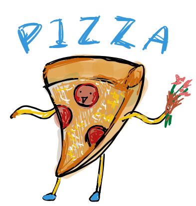
Note: A white noise sequence is 0-correlated and MA(1) is 1-correlated.

<u>Definition</u> The MA(q) process ↗ q-correlated.

$\{X_t\}$ is a moving-average process of order q if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\theta_1, \cdots \theta_q$ are constants.



2.2 Linear Process.

Every second order $E(X^2)$ weakly stationary process is either a linear process or can be transformed into one by subtracting a deterministic component.

<u>definition</u> $\{X_t\}$ is a linear process if

$$X_t = \sum_{j=-\infty}^{\infty} \varphi_j Z_{t-j}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\{\varphi_j\}$ is a sequence of constants with $\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty$

↘ ensure $E(X_t)$ converge

$$\begin{aligned} E(|X_t|) &\le E\left[\sum_{j=-\infty}^{\infty} |\varphi_j Z_{t-j}|\right] \\ &\le \sum_{j=-\infty}^{\infty} |\varphi_j| \, E|Z_{t-j}| \\ &\le \sum_{j=-\infty}^{\infty} |\varphi_j| \left[E(Z_{t-j}^2)\right]^{\frac{1}{2}} \\ &= \sum_{j=-\infty}^{\infty} |\varphi_j| \sigma = \sigma \sum_{j=-\infty}^{\infty} |\varphi_j| \\ &< \infty \qquad \text{if } \sum_{j=-\infty}^{\infty} |\varphi_j| < \infty \end{aligned}$$

Can view in terms of <u>backwards shift operator</u>

$X_t = \varphi(B) Z_t$ where $\varphi(B) = \sum_{j=-\infty}^{\infty} \varphi_j B^j$ recall $B^j Z_t = Z_{t-j}$

Note: A linear process is called a moving average $(MA(\infty))$ if

$$\varphi(B) = \sum_{j=0}^{\infty} \varphi_j B^j \quad \text{i.e. } \varphi_j = 0 \ \forall\, j < 0$$

Not allowed to have any dependence on the future.

e.g. $X_t = Z_t + 0.5 Z_{t-1}$ $\quad X_{t-1} = Z_{t-1} + 0.5 Z_{t-2}$

$\varphi_0 = 1, \ \varphi_1 = 0.5, \ \varphi_j = 0 \ \forall\, j \neq 0, 1$

Let Y_t be a stationary process with mean 0 and auto-covariance function $\gamma_Y(h)$.
If $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the time series

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(B) Y_t$$

is also a stationary sequence with mean 0 and auto-covariance function

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j)$$

where $Y_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, $\{Z_t\}$ is WN

Proof $\gamma_X(h) = E[X_t X_{t-h}]$ if $E(X_t) = 0$

$$= E\Big[\Big(\sum_{k=-\infty}^{\infty} \psi_k Y_{t-k}\Big)\Big(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j-h}\Big)\Big]$$

$$= E\Big[\sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_k Y_{t-k} \psi_j Y_{t-j-h}\Big]$$

$$= \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_k \psi_j E[\underbrace{Y_{t-k} \cdot Y_{t-j-h}}_{\gamma_Y(h+j-k)}]$$

$\to$ $\{X_t\}$ is the linear process

If $\{Y_t\}$ is a WN$(0,\sigma^2)$ process, then $\gamma_Y(l) = 0$ $\forall$ $l \neq 0$

$$\Rightarrow \gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-h} \sigma^2$$

e.g. AR(1)

For $\{X_t\}$ stationary, let $X_t = \phi X_{t-1} + Z_t$ where $\{Z_t\}$ is WN$(0,\sigma^2)$ and $\{X_t\}$ and $\{Z_t\}$ are uncorrelated for $s > t$

Define $\{X_t\}$ to be the solution to $X_t - \phi X_{t-1} = Z_t$

Consider $X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$ $\Rightarrow$ X_t is linear with $\psi_j = \phi^j$, $\psi_j = 0$ for $j < 0$

Note: $\sum_{-\infty}^{\infty} |\psi_j| < \infty$ iff $|\phi| < 1$

Claim $X_t - \phi X_{t-1} = Z_t$

$$\Big[\sum_{j=0}^{\infty} \phi^j Z_{t-j}\Big] - \phi\Big[\sum_{j=0}^{\infty} \phi^j Z_{t-1-j}\Big]$$

$$= \Big[\sum_{j=0}^{\infty} \phi^j Z_{t-j}\Big] - \Big[\sum_{j=0}^{\infty} \phi^{j+1} Z_{t-1-j}\Big]$$

$$= \Big[\sum_{j=0}^{\infty} \phi^j Z_{t-j}\Big] - \Big[\sum_{j=1}^{\infty} \phi^j Z_{t-j}\Big]$$

$$= \phi^0 Z_{t-0} = Z_t$$

$\{Z_t\}$ is stationary $\Rightarrow$ $\{X_t\}$ is stationary with mean 0 and ACVF

$$\gamma_X(h) = \sum_{i=0}^{\infty} \phi^i \phi^{i+h} \sigma^2$$
$$= \frac{\sigma^2 \phi^h}{1-\phi^2} \quad \text{for } h \geq 0$$

If $|\phi|>1$, then no stationary sequence exists that dependending on past

Use backwards shift operator

Let $\Phi(B) = 1-\phi B$ and $\pi(B) = \sum_{i=0}^{\infty} \phi^i B^i$

then $\varphi(B) = \Phi(B)\pi(B) = (1-\phi B)\sum_{i=0}^{\infty}\phi^i B^i = \sum_{i=0}^{\infty}\phi^i B^i - \sum_{i=0}^{\infty}\phi^{i+1}B^{i+1} = \phi^0 B^0 = 1$

$$Z_t = X_t - \phi X_{t-1} = (1-\phi B)X_t = \Phi(B)X_t$$
$$\pi(B)Z_t = \pi(B)\Phi(B)X_t = X_t$$
$$\Rightarrow X_t = \sum_{i=0}^{\infty}\phi^i B^i Z_t = \sum_{i=0}^{\infty}\phi^i Z_{t-i}$$

2.4 Properties of the Sample Mean and Autocorrelation Function.

① Estimation of μ

$$\hat{\mu} = \bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

- It is unbiased because $E(\bar{X}_n) = \dfrac{E(X_1) + \cdots + E(X_n)}{n} = \mu$
- The mean squared error of $\bar{X}_n$ is

$$E[(\bar{X}_n - \mu)^2] = MSE(\bar{X}_n) = Var(\bar{X}_n)$$
$$Var(\bar{X}_n) = Var\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) = \frac{1}{n^2} Var(X_1 + \cdots + X_n)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma(i-j)$$

$\{ 0 \leq |i-j| \leq n$, $\gamma(h) = \gamma(-h)$ $\quad$ $n+1$ values of $|i-j|$

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

number of possible observed lags.

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

- If $\gamma(h) \to 0$ as $n \to \infty$, then $Var(\bar{X}_n) = MSE(\bar{X}_n) \to 0$ as $n \to \infty$
- If $\sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty$, then $\lim_{n\to\infty} n\,Var(\bar{X}_n) = \sum_{h=-\infty}^{\infty}|\gamma(h)|$

If $\{X_t\}$ are also Gaussian, then $\bar{X}_n \sim N(\mu, \frac{1}{n} \sum_{|h|<\infty} (1-\frac{|h|}{n}) \gamma(h))$

Testing and Interval Estimation.

For many time series, in particular for linear and ARMA models, $\bar{X}_n$ is approximately normal with mean μ and variance $\frac{1}{n} \sum_{|h|<\infty} \gamma(h)$ for large n. An approximate CI for μ is then

$$\bar{X}_n \pm Z_{\alpha/2} \frac{\sqrt{v}}{\sqrt{n}} \quad \text{where } v = \sum_{|h|<\infty} \gamma(h)$$

generally unkown

Plug in $\hat{v} = \sum_{|h|<\sqrt{n}} \hat{\gamma}(h)$ $\quad \hat{v} = \sum_{|h|<\sqrt{n}} (1-\frac{|h|}{\sqrt{n}}) \hat{\gamma}(h)$?

e.g. AR(1) Let $\{X_t\}$ be stationary

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

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

$$= (1+2\phi \sum_{h=1}^{\infty} \phi^{h-1}) \frac{\sigma^2}{1-\phi^2}$$

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

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

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

$\Rightarrow$ 95% CI for μ

$$\bar{X}_n \pm \frac{1.96}{\sqrt{n}} \frac{\sigma}{|1-\phi|}$$

Both are unkown in practice, they must be replaced by estimated values.

② Estimation of $\gamma(\cdot)$ and $\rho(\cdot)$

Sample ACVF and ACF are defined by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n)$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

However, both estimators $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are biased in finite sequence.

The sample ACVF has the desirable property that for each $k \geq 1$, the k-dimensional sample covariance matrix

$$\Gamma_k = \begin{pmatrix} \hat{\gamma}(0) & \cdots & \hat{\gamma}(k-1) \\ \hat{\gamma}(1) & \ddots & \\ \vdots & & \ddots \\ \hat{\gamma}(k-1) & & \hat{\gamma}(0) \end{pmatrix}$$

is nonnegative definite.

Without further information beyond the observed data $X_1 \cdots X_n$, it is impossible to give reasonable estimates of $\gamma(h)$ and $\rho(h)$ for $h \geq n$. Even for h slightly smaller than n, the estimates $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are unreliable, since there are so few pairs (X_{t+h}, X_t) available (only one if $h = n-1$). A useful guide is provided by Jenken's rule & theorem, who suggests that n should be at least 50 and $h \leq \frac{n}{4}$.

For large sample without large lag, we can approximate the distribution

$$f(\hat{\rho}(1) \cdots \hat{\rho}(k)) \quad \text{by} \quad \hat{\rho} \dot{\sim} MVN(\underset{\sim}{\rho}, \tfrac{1}{n} W)$$

where W is a $k \times k$ covariance matrix with elements computed by a simplication of Bartlett's formula.

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(k)\rho(i)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(k)\rho(j)\}$$

e.g. iid $\{X_t\}$ $\Rightarrow$ $\rho(0) = 1$, $\rho(h) = 0$ $|h| \geq 1$

$$\text{So } w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(h)\rho(i)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(h)\rho(j)\}$$

$$= \sum_{k=1}^{\infty} \rho(k-i)\rho(k-j)$$

$$\Rightarrow w_{ii} = \rho(0)\rho(0) = 1$$

$$w_{ij} = 0$$

Therefore, $\hat{\rho}(h) \dot{\sim} N(0, \frac{1}{n})$ as $W = I$

e.g. MA(1)

$X_t = Z_t + \theta Z_{t-1}$, $Z_t \sim WN(0, \sigma^2)$

Then $X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k} \Rightarrow \psi_k = \begin{cases} 1 \text{ for } k=0 \\ \theta \text{ for } k=1 \\ 0 \text{ for } k \neq \{0,1\} \end{cases}$

So $\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-h} \sigma^2$ $\gamma_Z(0)$

for $h=0$ $\gamma(0) = \sum_{j=-\infty}^{\infty} \psi_j^2 \sigma^2 = (1+\theta^2)\sigma^2$. $j=0$ or 1

$h=1$ $\gamma(1) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-1} \sigma^2 = (\theta \cdot 1)\sigma^2 = \theta\sigma^2$ for $h \geq 2 = 0$ $j=1$

$h=-1$ $\gamma(-1) = \theta\sigma^2$ $j=0$

$|h|>1$ $\psi_j \psi_{j-h} = 0$

$\Rightarrow \rho(h) = \begin{cases} 1 & h=0 \\ \frac{\theta}{1+\theta^2} & h=\pm 1 \\ 0 & |h|>1 \end{cases}$

Therefore,

$w_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(k)\rho(i)][\rho(k+j) + \rho(k-j) - 2\rho(k)\rho(j)]$

for $i=j$, $w_{ii} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(k)\rho(i)\}^2$

- $w_{11} = \sum_{k=1}^{\infty} [\rho(k+1) + \rho(k-1) - 2\rho(k)\rho(1)]^2$

Strategy: Identity $\rho(i)$ where $\rho(i)$ are not zero

$[\rho(0) - 2\rho(1)\rho(1)]^2 + [\rho(1)]^2$

$= 1 - 3\rho(1)^2 + 4\rho(1)^4$

- for $i>1$

$w_{ii} = \sum_{k=1}^{\infty} \{\underbrace{\rho(k+i)}_{\text{always } 0} + \underbrace{\rho(k-i)}_{\substack{\text{nonzero} \\ \text{for } k=i,i+1,i-1}} - 2\underbrace{\rho(k)\rho(i)}_{\text{always zero}}\}^2$

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

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

Thus, $\hat{\rho}(1) \dot{\sim} N\left(\frac{\theta}{1+\theta^2}, \frac{1}{n}(1-3\rho(1)^2 + 4\rho(1)^4)\right)$

2.5 Forecasting Stationary Time Series.

Goal: Find linear combination of $X_1, \cdots X_n$ that forecasts X_{n+h} with minimum MSE.

Best linear predictor of X_{n+h}

$$P_n X_{n+h} = a_0 + a_1 X_n + a_2 X_{n-1} + \cdots + a_n X_1$$

Find $a_0, a_1 \cdots a_n$ to minimize

$$S(a_0, a_1 \cdots a_n) = E[(X_{n+h} - (a_0 + a_1 X_n + \cdots + a_n X_1))^2]$$

S is quadratic, bounded below by zero. Thus, $\exists$ at least one solution to $\frac{\partial}{a_j} S(a_0, a_1 \cdots a_n) = 0$ for $j = 0, \cdots n$

Assume interchang $\frac{\partial}{\partial a_j}$ and $E(\cdot)$ safely

① $E[X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i}] = 0$

$$\Rightarrow a_0 = E(X_{n+h}) - \sum_{i=1}^n a_i E(X_{n+1-i})$$
$$= \mu - \mu \sum_{i=1}^n a_i \quad \text{if } \{X_n\} \text{ are stationary}$$
$$= (1 - \sum_{i=1}^n a_i)\mu$$

② $E[(X_{n+h} - (1 - \sum_{i=1}^n a_i)\mu - \sum_{i=1}^n a_i X_{n+1-i}) X_{n+1-j}] = 0$

$$\Rightarrow \underbrace{E[(X_{n+h} - \mu) X_{n+1-j}]}_{\gamma(h+j-1)} = E[(\sum_{i=1}^n a_i (X_{n+1-i} - \mu)) X_{n+1-j}]$$

$\gamma(n+1-h)$

b.c. $E(X_{n+1-h} - \mu) = 0$, then use $Cov(X,Y) = E(XY) - E(X)E(Y)$

$$\downarrow$$

$$\sum_{i=1}^n a_i E[(X_{n+1-i} - \mu) X_{n+1-j}]$$

$$\sum_{i=1}^n a_i \gamma(i-j)$$

$$\Rightarrow \gamma(n+1-h) = \sum_{i=1}^n a_i \gamma(i-j)$$

$$\underset{n\times n}{\Gamma_n}\, \underset{\sim}{a_n} = \underset{n\times 1}{\underset{\sim}{\gamma_n}(h)} \quad \text{where } (\Gamma_n)_{ij} = \gamma(i-j)$$

$$\gamma_n(h) = (\gamma(h), \gamma(h+1) \cdots \gamma(h+n-1))$$

$$\underset{\sim}{a_n}(h) = (a_1, \cdots a_n)$$

Therefore, $P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)$ where

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

Obviously, $E[X_{n+h} - (\mu + \sum_{i=1}^{n} a_i (X_{n+h} - \mu))] = 0$

MSE: $E[(X_{n+h} - P_n X_{n+h})^2]$

$$= E[(X_{n+h} - (\mu + \sum_{i=1}^{n} a_i (X_{n+h} - \mu)))^2]$$

$$= E[(X_{n+h} - \mu) - (\sum_{i=1}^{n} a_i (X_{n+1-i} - \mu))^2]$$

$$= E[(X_{n+h} - \mu)^2] - 2E[(X_{n+h} - \mu)(\sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)] + E[\sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)]^2$$

$$= \gamma(0) - 2\sum_{i=1}^{n} a_i \gamma(h+i-1) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma(i-j) a_j$$

$$= \gamma(0) - 2\sum_{i=1}^{n} a_i \gamma(h+i-1) + \sum_{i=1}^{n} a_i (\sum_{j=1}^{n} \gamma(i-j) a_j)$$

$$= \gamma(0) - 2\underset{\sim}{a_n} \gamma_n(h) + \underset{\sim}{a_n} \Gamma \underset{\sim}{a_n}$$

$$= \gamma(0) - \underset{\sim}{a_n} \gamma_n(h)$$

e.g. One-step prediction of AR(1)

$X_t = \phi X_{t-1} + Z_t \quad t = 0, \pm 1 \cdots$ where $|\phi| < 1$ and $\{Z_t\} \sim WN(0, \sigma^2)$.

$\Rightarrow P_n(X_{n+1}) = \underset{\sim}{a_n}^T \underset{\sim}{X_n}$ where $\underset{\sim}{X_n} = (X_n, \cdots X_1)$

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

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \\ & & \ddots & & \\ & & & & \\ & & & & \gamma(1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \vdots \\ \gamma(n) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & & & \\ \phi^2 & \phi & 1 & \phi & & \\ \vdots & & \phi & 1 & \ddots & \\ & & & & & \\ \phi^{n-1} & & & & & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \vdots \\ \phi^n \end{bmatrix}$$

non-negative definite

$\Rightarrow$ unique solution

Try $a_1 = \phi_1$, $a_k = 0$ for $k = 2 \cdots n$

$\rightarrow$ Solution:

$\underset{\sim}{a_n} = (\phi, 0, 0 \cdots 0)$ and $P_n X_{n+1} = \underset{\sim}{a_n}^T \underset{\sim}{X_n} = \phi X_n$

Note: This solution doesn't work for MA(1).

$$
\begin{aligned}
E[(X_{n+1} - P_n X_{n+1})^2] &= \gamma(0) - \underset{\sim}{a_n}^T \underset{\sim}{\gamma}(1) = \frac{\sigma^2}{1-\phi^2} - \phi\gamma(1) \\
&= \frac{\sigma^2}{1-\phi^2} - \phi \frac{\sigma^2\phi}{1-\phi^2} \\
&= \frac{\sigma^2(1-\phi^2)}{1-\phi^2} \\
&= \sigma^2.
\end{aligned}
$$

① Prediction of Second Order Random variable.

Suppose now that Y and $W_n \cdots W_1$ are any random variables with finite second moments and that the mean $\mu = E(Y)$, $\mu_i = E(W_i)$ and covariances $Cov(Y, W_i)$, $Cov(W_i, W_j)$ $\forall i, j$ are all known.

<u>Notation</u>

$\underset{\sim}{W} = (W_n, \cdots W_1)^T$

$\underset{\sim}{\mu_W} = (\mu_n, \cdots \mu_1)^T$

$\underset{\sim}{\gamma} = Cov(Y, \underset{\sim}{W}) = (Cov(Y, W_n) \cdots Cov(Y, W_1))^T$

$\Gamma = Cov(\underset{\sim}{W}, \underset{\sim}{W}) = [Cov(W_{n+1-i}, W_{n+1-j})]_{i,j=1}^{n}$

Then by the same method from before, we suggest that

$P(Y | \underset{\sim}{W}) = \mu_Y + \underset{\sim}{a}^T(\underset{\sim}{W} - \underset{\sim}{\mu_W})$ where

$\underset{\sim}{a} = (a_1 \cdots a_n)$ is any solution of $\Gamma \underset{\sim}{a} = \underset{\sim}{\gamma}$

Recall: $(X^T X)\beta = X^T Y$

$$\left[\frac{\sum_{i=1}^{n} x_{ij} x_{ik}}{n} \right] \qquad \left[\frac{\sum_{i=1}^{n} x_{ij} y_i}{n} \right]$$

$\frac{1}{n}$ $\qquad$ $\frac{1}{n}$

The mean squared error of the predictor is

$E[(Y - P(Y|\underset{\sim}{W}))^2] = Var(Y) - \underset{\sim}{a}^T \underset{\sim}{\gamma}$

e.g. Estimation of missing value (AR(1))

Assume that we obtain X_1 and X_3, but not X_2

Goal: Find linear predictor for X_2

Let $Y = X_2$, $W = (X_1, X_3)$

$$\Gamma = \begin{pmatrix} \frac{\sigma^2}{1-\phi^2} & \frac{\phi^2\sigma^2}{1-\phi^2} \\ \frac{\phi^2\sigma^2}{1-\phi^2} & \frac{\sigma^2}{1-\phi^2} \end{pmatrix}$$

$$\gamma = \begin{pmatrix} \frac{\sigma^2}{1-\phi^2}\phi \\ \frac{\sigma^2}{1-\phi^2}\phi \end{pmatrix} \begin{array}{l} \leftarrow Cov(X_2, X_1) \\ \leftarrow Cov(X_2, X_3) \end{array}$$

$$\Rightarrow \quad \Gamma \underset{\sim}{a} = \underset{\sim}{\gamma}$$

$$\begin{pmatrix} 1 & \phi^2 \\ \phi^2 & 1 \end{pmatrix} \underset{\sim}{a} = \begin{pmatrix} \phi \\ \phi \end{pmatrix}$$

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

$$\Rightarrow \quad P(X_2 \mid X_1, X_3) = \frac{\phi}{1+\phi^2} X_1 + \frac{\phi}{1+\phi^2} X_3$$

$$= \frac{\phi}{1+\phi^2}(X_1 + X_3) + \left(1 - \frac{\phi}{1+\phi^2}\right) \cdot 0$$

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

② The Prediction Operator $P(\cdot \mid W)$

Properties of the Prediction Operator $P(\cdot \mid W)$

Suppose that $E(U^2) < \infty$, $E(V^2) < \infty$, $\Gamma = Cov(\underset{\sim}{W}, \underset{\sim}{W})$ and $\beta, \alpha_1, \cdots \alpha_n$ are constants.

① $P(U \mid \underset{\sim}{W}) = E(U) + a^T(\underset{\sim}{W} - E\underset{\sim}{W})$ where $\Gamma \underset{\sim}{a} = Cov(U, \underset{\sim}{W})$

② $E[(U - P(U \mid \underset{\sim}{W}))\underset{\sim}{W}] = 0$ and $E[U - P(U \mid \underset{\sim}{W})] = 0$.

③ $E[(U - P(U \mid \underset{\sim}{W}))^2] = Var(U) - a^T Cov(U, \underset{\sim}{W})$

④ $P(\alpha_1 U + \alpha_2 V + \beta \mid \underset{\sim}{W}) = \alpha_1 P(U \mid \underset{\sim}{W}) + \alpha_2 P(V \mid \underset{\sim}{W}) + \beta$

⑤ $P(\sum_{i=1}^{n} \alpha_i W_i + \beta \mid \underset{\sim}{W}) = \sum_{i=1}^{n} \alpha_i W_i + \beta$

⑥ $P(U \mid \underset{\sim}{W}) = E(U)$ if $Cov(U, \underset{\sim}{W}) = \underset{\sim}{0}$

⑦ $P(U \mid \underset{\sim}{W}) = P(P(U \mid \underset{\sim}{W}, \underset{\sim}{V}) \mid \underset{\sim}{W})$ if V is a random vector s.t. the components of $E(VV^T)$ are all finite.

e.g. AR(p)

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

We can apply the prediction operator P_n to each side of the defining equations.

using ④⑤⑥, we get

$$P_n X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p}$$

Remark 3. In general, if $\{Y_t\}$ is a stationary time series with mean μ and if $\{X_t\}$ is the zero-mean series defined by $X_t = Y_t - \mu$, then since the collection of all linear combinations of $1, Y_n, \ldots, Y_1$ is the same as the collection of all linear combinations of $1, X_n, \ldots, X_1$, the linear predictor of any random variable W in terms of $1, Y_n, \ldots, Y_1$ is the same as the linear predictor in terms of $1, X_n, \ldots, X_1$. Denoting this predictor by $P_n W$ and applying P_n to the equation $Y_{n+h} = X_{n+h} + \mu$ gives

$$P_n Y_{n+h} = \mu + P_n X_{n+h}. \tag{2.5.19}$$

Thus the best linear predictor of Y_{n+h} can be determined by finding the best linear predictor of X_{n+h} and then adding μ. Note from (2.5.8) that since $E(X_t) = 0$, $P_n X_{n+h}$ is the same as the best linear predictor of X_{n+h} in terms of $X_n, \ldots, X_1$ only. □

e.g. AR(1) with Nonzero mean

$\{Y_t\}$ is said to be an AR(1) process with mean μ if $\{X_t = Y_t - \mu\}$ is a zero-mean AR(1) process. Let $Y_t = X_t + \mu$

$$\Rightarrow P_n Y_{n+h} = \mu + \phi^h (Y_n - \mu)$$

Assume $\{X_t\}$ is a stationary process, with mean 0 and covariance $\gamma(\cdot)$.
Can solve for $\underset{\sim}{a}$ to determine $P_n X_{n+h}$ in terms of $\{X_n \cdots X_1\}$
However, for large n, inverting Γ is NOT fun

$\Rightarrow$ We could use linearity of P_n to do recursive prediction of $P_{n+1} X_{n+h}$ from $P_n X_{n+1}$

If Γ_n is nonsingular, then

$$P_n X_{n+1} = \underset{\sim}{\phi_n}^T \underset{\sim}{X_n} = \phi_1 X_n + \cdots + \phi_n X_1$$

③ Durbin-Levinson Algorithm

Assumption

$E(X)=0$, $\{X_t\}$ stationary

Algorithm

Compute coefficients $\phi_{n1}, \cdots \phi_{nn}$ at step n of the algorithm

At step n, best linear predictor is of the form $P_n X_{n+1} = \sum_{j=1}^{n} \phi_{nj} X_{n-j+1}$

Predict X_{n+1}.

At the end, $P_N X_{N+1} = \sum_{j=1}^{N} a_j X_{N-j+1}$

The Durbin–Levinson Algorithm:
The coefficients $\phi_{n1}, \ldots, \phi_{nn}$ can be computed recursively from the equations

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] v_{n-1}^{-1}, \tag{2.5.20}$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \tag{2.5.21}$$

and

$$v_n = v_{n-1}\left[1 - \phi_{nn}^2\right], \tag{2.5.22}$$

where $\phi_{11} = \gamma(1)/\gamma(0)$ and $v_0 = \gamma(0)$.

Proof Now let $R_n = \frac{1}{\gamma(0)} \Gamma_n$ $\qquad R_n = \begin{pmatrix} 1 & \rho(1) & \\ & \ddots & \\ & & 1 \end{pmatrix}$

Notation $\rho_{1:n} = (\rho(1) \cdots \rho(n))^T$

$\rho_{n:1} = \rho_{1:n}^{R} = (\rho(n) \cdots \rho(1))^T$ Ⓡ reverse

If $\Gamma_n \underset{\sim}{\phi}_n = \underset{\sim}{\gamma}(n)$, then $R_n \underset{\sim}{\phi}_n = \underset{\sim}{\rho}_{1:n}(n)$ dividing by $\gamma(0)$

① base case

$R_1 \phi_{11} = \rho(1)$ because $R_1 = \rho(0) = 1$

② inductive step

Assume $R_k \underset{\sim}{\phi}_k = \rho_{1:k}$ holds $\qquad \underset{\sim}{\phi}_k = (\phi_{k1} \cdots \phi_{kk})^T$

Need to show $R_{k+1} \underset{\sim}{\phi}_{k+1} = \rho_{1:k+1}$

Now also define $\phi_{n:1} = \phi_{1:n}^{R} = (\phi_{nn} \cdots \phi_{n1})$

$R_{k+1}\boldsymbol{\phi}_{k+1}$

$$= \begin{bmatrix} R_k & \boldsymbol{\rho}_{k:1} \\ \boldsymbol{\rho}^T_{k:1} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi}_k - \phi_{k+1,k+1}\boldsymbol{\phi}^R_k \\ \phi_{k+1,k+1} \end{bmatrix}$$

$1 \rightarrow \rho(0)$

$$= \begin{bmatrix} R_k(\boldsymbol{\phi}_k - \phi_{k+1,k+1}\boldsymbol{\phi}^R_k) + \boldsymbol{\rho}_{k:1}\phi_{k+1}\phi_{k+1} \\ \boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}_k - \boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}^R_k\phi_{k+1,k+1} + \phi_{k+1,k+1} \end{bmatrix}$$

$$= \begin{bmatrix} R_k\boldsymbol{\phi}_k - R_k\phi_{k+1}\phi_{k+1}\boldsymbol{\phi}^R_k + \boldsymbol{\rho}_{k:1}\phi_{k+1}\phi_{k+1} \\ \boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}_k + \phi_{k+1,k+1}(1-\boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}^R_k) \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\rho}_{1:k} + \phi_{k+1,k+1}(\boldsymbol{\rho}_{k:1} - \underbrace{R_k\boldsymbol{\phi}^R_k}_{\rightarrow\,\boldsymbol{\rho}_{k:1}}) \\ \boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}_k + \phi_{k+1,k+1}(1-\boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}^R_k) \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\rho}_{1:k} \\ * \end{bmatrix}$$

$$\boxed{\begin{aligned} \phi_{k+1,k+1} &= \frac{(\gamma(k+1) - \boldsymbol{\gamma}^T_{k:1}\boldsymbol{\phi}_k)}{v_k} \\ \text{where}\quad v_n &= E((X_{n+1} - \textstyle\sum_{j=1}^{k}\phi_k X_{k-j+1})^2) \\ &= \gamma(0) - \boldsymbol{\phi}_k^T\boldsymbol{\gamma}_{k:1} \\ &= \gamma(0)(1 - \boldsymbol{\phi}_k^T\boldsymbol{\rho}_{k:1}) \\ \phi_{k+1,k+1} &= \frac{(\gamma(k+1) - \boldsymbol{\gamma}^T_{k:1}\boldsymbol{\phi}_k)}{\gamma(0)(1-\boldsymbol{\phi}^T_k\boldsymbol{\rho}_{k:1})} \end{aligned}}$$

Then, $*$ is equal to

$$\boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}_k + \frac{(\gamma(k+1) - \boldsymbol{\gamma}^T_{k:1}\boldsymbol{\phi}_k)}{\gamma(0)(1-\boldsymbol{\phi}^T_k\boldsymbol{\rho}_{k:1})}(1-\boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}^R_k)$$

equal (reverse of inner product)

$$= \boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}_k + \frac{\gamma(k+1)}{\gamma(0)} - \boldsymbol{\rho}^T_{k:1}\boldsymbol{\phi}_k$$

$$= \rho(k+1)$$

$$\begin{aligned} v_n &= E[(X_{n+1} - \textstyle\sum_{j=1}^{n}\phi_{nj}X_{n-j+1})^2] \\ &= \gamma(0) - \boldsymbol{\phi}_n^T\boldsymbol{\gamma}_n \\ &= \gamma(0) - \boldsymbol{\phi}^T_{n-1}\boldsymbol{\gamma}_{n-1} + \phi_{nn}\boldsymbol{\phi}^{RT}_{n-1}\boldsymbol{\gamma}_{n-1} - \phi_{nn}\gamma(n) \\ &= \gamma(0) - \boldsymbol{\phi}^T_{n-1}\boldsymbol{\gamma}_{n-1} - \phi_{nn}(\gamma(n) - \boldsymbol{\phi}^{RT}_{n-1}\boldsymbol{\gamma}_{n-1}) \\ &= v_{n-1} - \phi_{nn}(\phi_{nn}v_{n-1}) \\ &= v_{n-1}(1-\phi^2_{nn}) \end{aligned}$$

recursion step.

④ Innovation Algorithm

$\{X_t\}$ is zero-mean with $E(X^2)<\infty$ and $E(X_i X_j) = k(i,j)$

Denote the best one-step linear predictors and their MSE by

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

and

$$v_n = E[(X_{n+1} - P_n X_{n+1})^2]$$

We shall introduce the innovations, or one-step prediction error

$$U_n = X_n - \hat{X}_n$$

In terms of the vector $\underset{\sim}{U_n} = (U_1 \dots U_n)^T$ and $\underset{\sim}{X_n} = (X_1 \dots X_n)^T$, we then have

$$\underset{\sim}{U_n} = A_n \underset{\sim}{X_n} \quad \text{where} \quad A_n = \begin{pmatrix} 1 & & & 0 \\ a_{11} & 1 & & \\ a_{22} & a_{21} & 1 & \\ \dots & \dots & \dots & \end{pmatrix}$$

i.e. $U_1 = X_1$ for mean-zero sequence

$U_2 = X_2 + a_{11} X_1$

$U_3 = X_3 + (a_{21} X_2 + a_{22} X_1)$

$\dots$

Note if $\{X_t\}$ is stationary, then $a_{ij} = -a_j$ where a_j from $\Gamma_n \underset{\sim}{a} = \underset{\sim}{\gamma_n}$

This implies that A_n is nonsingular with inverse $C_n = A_n^{-1}$ of the form

$$C_n = \begin{pmatrix} 1 & & & 0 \\ \theta_{11} & 1 & & \\ \theta_{22} & \theta_{21} & 1 & \\ \dots & \dots & & \\ \dots & \dots & & \end{pmatrix} \qquad \Theta_n = \begin{pmatrix} 0 & & & 0 \\ \theta_{11} & 0 & & \\ \theta_{22} & \theta_{21} & 0 & \\ \dots & & & \ddots \\ \dots & & & \end{pmatrix}$$

The vector of one-step predictors $\underset{\sim}{\hat{X}_n} = (X_1, P_1 X_2, \dots P_n X_{n+1})^T$ can therefore be expressed as

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

$$\Rightarrow \quad \hat{X}_{n+1} = \begin{cases} 0 & n=0 \\ \sum_{j=1}^{n} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) & n=1,2\dots \end{cases}$$

The Innovations Algorithm:

The coefficients $\theta_{n1}, \ldots, \theta_{nn}$ can be computed recursively from the equations

$$v_0 = \kappa(1, 1),$$

$$\theta_{n,n-k} = v_k^{-1}\left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right), \quad 0 \le k < n,$$

and

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j.$$

(It is a trivial matter to solve first for v_0, then successively for $\theta_{11}, v_1; \theta_{22}, \theta_{21}, v_2; \theta_{33}, \theta_{32}, \theta_{31}, v_3; \ldots$.)