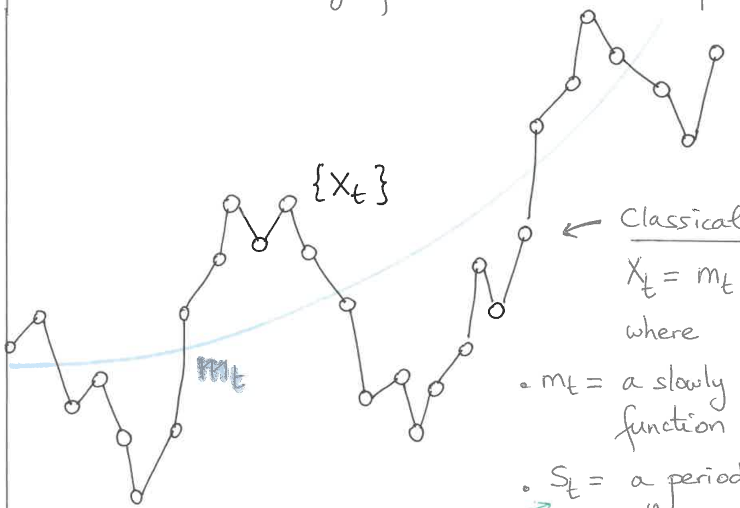
# TS: ARIMA PROCESSES

## I_ STATIONARY TIME SERIES

A time series model specifies the joint distribution of a sequence $\{X_t\}$ of random variables:

$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$ $\forall n$ and $x_1, \dots, x_n$

↳ In practice, it is a good idea to plot the time series; and to be looking for trends, seasonal components, outliers...




← Classical decomposition:

$X_t = m_t + S_t + Y_t$

where

- $m_t$ = a slowly varying function aka TREND.
- $S_t$ = a periodic function, with period $d$
- $Y_t$ = a zero-mean process.

seasonal adjustment

x Goal: Remove trend & seasonality to get STATIONARY (to be defined shortly) RESIDUALS.

↖ Hopefully with some dependence structure that we can take advantage of. Fit a model to the residuals & forecast.

- Ex of time series

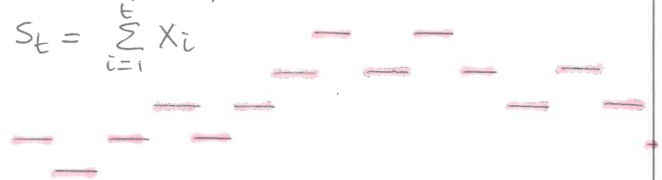
2

(i) iid noise $\{X_t\}$ with $EX_t = 0$, $EX_t^2 = \sigma^2 < +\infty$, iid

(ii) white noise $\{X_t\}$ = sequence of uncorrelated RVs with zero mean and variance $\sigma^2$.

We write $X_t \sim WN(0, \sigma^2)$

(iii) random walk $S_t = \sum_{i=1}^{t} X_i$



Elimination of Trend & Seasonal Component.

There are two main approaches

(a) estimate $m_t$ and $S_t$

(b) apply differencing operator to $X_t$ (Box & Jenkins '76)

x Case I: $X_t = m_t + Y_t$ (trend only)

(a) Non-parametric estimate using a moving average

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-q}, \text{ or } \hat{m}_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1} \text{ (exp. smoothing)}$$

↖ q = window size

(b) define the lag-1 difference operator $\nabla X_t := X_t - X_{t-1}$

"think first derivative"

$\nabla X_t = (1-B) X_t$, where $B$ = backward shift operator

$B^j X_t = X_{t-j}$

then
$$\begin{aligned}\nabla^2 X_t &:= \nabla(\nabla X_t)\\ &= \nabla(X_t - X_{t-1})\\ &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})\\ &= X_t - 2X_{t-1} + X_{t-2} = (1-2B+B^2)X_t\\ &= (1-B)(1-B)X_t\end{aligned}$$

polynomials in $B$ & $\nabla$ are manipulated as poly. of real variable

Why considering the difference operator?

↳ If $m_t = \beta_0 + \beta_1 t$, then $\nabla m_t = \beta_1 =$ constant

↳ Likewise, if $m_t = \sum_{j=0}^{k} \beta_j t^j$, then $\nabla^k m_t = k!\,\beta_k$ = constant

the trend component is removed

x Case II: $X_t = m_t + S_t + Y_t$ (trend + seasonal component)

where
- $S_{t+d} = S_t$, $d$ known
- $\sum_{j=1}^{d} S_j = 0$ "zero mean"
- $E\,Y_t = 0$

(a) Proceed as before, with $d = 2q+1$ (if $d$ is odd) & compute the average $w_k$ of deviations $x_{k+jd} - \hat{m}_{k+jd}$, for each $k = 1, \dots, d$. We estimate the seasonal component $s_k$ as

$$\hat{s}_k = w_k - \frac{1}{d}\sum_{i=1}^{d} w_i \quad \text{(zero mean)}.$$

(b) Define the lag-d operator

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)\,X_t \neq \nabla^d X_t = (1-B)^d X_t$$

Applying $\nabla_d$ to $X_t$ yields

$$\nabla_d X_t = \boxed{m_t - m_{t-d}} + \boxed{Y_t - Y_{t-d}}$$

trend component: can be eliminated using powers of $\nabla$ — residual.

Remark: Variance-stabilizing transformations can be applied prior to trend & seasonality removal, such as $\log$, $\sqrt{\ }$, ... see (Box and Cox (1964))



## I.1. Stationarity

A process is stationary if it has "similar" properties as the time-shifted series $X_{t+h}$, $\forall h \in \mathbb{Z}$.

Formally, $\{X_t\}$ is strictly stationary if $\forall k, t_1, \dots, t_k, h$ and $x_1, \dots, x_k$,

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

shifting the time axis does not change the distribution

→ $\{X_t\}$ is (WEAKLY) STATIONARY if $\mu_X(t) := \mathbb{E}\,X_t$ (mean function) is independent of $t$, and, with

$$\gamma_X(r,s) := \mathrm{Cov}(X_r, X_s) = \mathbb{E}\{(X_r - \mathbb{E}X_r)(X_s - \mathbb{E}X_s)\},$$

(autocovariance function) (ACF)

$\forall h$, $\gamma_X(t+h, t)$ is independent of $t$. Second order properties only.

And we write $\gamma_X(h) := \gamma_X(h, 0)$.

The AUTOCORRELATION FUNCTION (ACF) of $\{X_t\}$ is defined as

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Corr}(X_{t+h}, X_t)$$

- Ex: (i) iid noise $\gamma_X(t+h, t) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{otherwise} \end{cases} = \gamma_X(h, 0) \quad \forall t$

  $E\,X_t = 0$ ← independent of $t$

  ⇒ $\{X_t\}$ is stationary & similarly for any WN sequence (i.e. zero mean, uncorrelated)

(ii) <u>Random walk</u> $S_t = \sum_{i=1}^{t} X_i$ .

(5)

- $\mathbb{E} S_t = 0 \leftarrow$ independent of $t$
- $\mathbb{E} S_t^2 = t\sigma^2$
- $\gamma_S(t+h, t) = \mathrm{Cov}(S_{t+h}, S_t)$

$$= \mathrm{Cov}\left(S_t + \sum_{j=1}^{h} X_{t+j},\ S_t\right)$$

$$= \mathrm{Cov}(S_t, S_t) = t\sigma^2$$

$\uparrow$ not independent of $t$.

$\Rightarrow \{S_t\}$ is not stationary.

(iii) <u>MA(1) (Moving Average)</u>: $X_t = Z_t + \theta Z_{t-1}$, where $Z_t \sim WN(0, \sigma^2)$

- $\mathbb{E} X_t = 0$
- $\mathbb{E} X_t^2 = \underbrace{\mathbb{E} Z_t^2}_{\sigma^2} + 2\theta\, \mathbb{E} Z_t Z_{t-1} + \theta^2 \underbrace{\mathbb{E} Z_{t-1}^2}_{\sigma^2} = \sigma^2(1+\theta^2)$

$$= \gamma_X(0)$$

- $\gamma_X(1) = \mathrm{Cov}(X_t, X_{t+1})$

$$= \mathbb{E}\,(Z_t + \theta Z_{t-1})(Z_{t+1} + \theta Z_t)$$

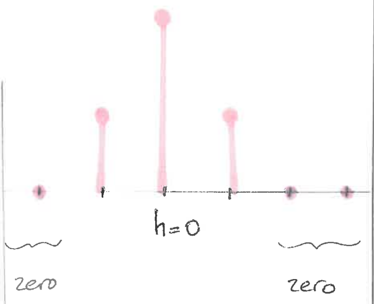
$$= \cancel{\mathbb{E}\, Z_t Z_{t+1}} + \theta \underbrace{\mathbb{E}\, Z_t^2}_{\sigma^2} + \cancel{\theta\, \mathbb{E}\, Z_{t-1} Z_{t+1}} + \cancel{\theta^2 \mathbb{E}\, Z_{t-1} Z_t}$$

$$= \theta\sigma^2$$

$=$ independent of $t$

- $\gamma_X(h) = 0$ for $h \geq 2$

$$\gamma_X(h) = \begin{cases} \sigma^2(1+\theta^2) & \text{if } h=0 \\ \theta\sigma^2 & \text{if } h=1 \\ 0 & \text{if } h \geq 2 \end{cases}$$




(iv) <u>AR(1) (Auto Regressive)</u> $X_t = \phi X_{t-1} + Z_t$,

(6)

where $Z_t \sim WN(0, \sigma^2)$, $|\phi| < 1$ (we will see why shortly)

& $Z_t$ is uncorrelated with $X_s$, $\forall s < t$ & stationary

- $\mathbb{E} X_t = \phi\, \mathbb{E} X_{t-1} = 0$ (from stationarity)
- $\gamma_X(h) \Rightarrow$ $(h>0)$

$$X_t = \phi X_{t-1} + Z_t \quad \times X_{t-h}$$

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

$\mathbb{E}(\cdot)$

$$\gamma_X(h) = \phi\, \gamma_X(h-1) + 0$$

$$= \dots = \phi^h \gamma_X(0)$$

Since $\gamma_X(h) = \gamma_X(-h)$, we have $\gamma_X(h) = \phi^{|h|} \gamma_X(0)$,

and $\rho_X(h) = \phi^{|h|}$, $\forall h$

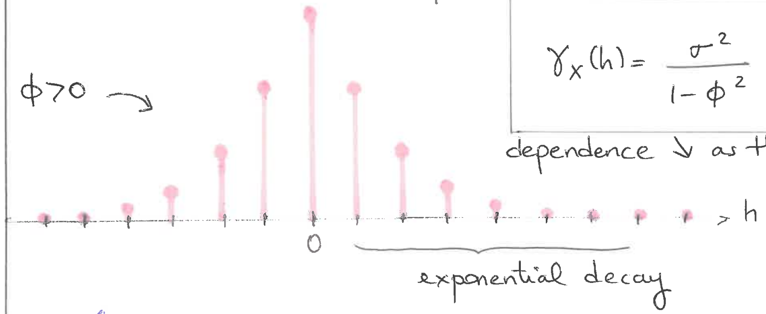
- Moreover, $\gamma_X(0) = \mathrm{Cov}(X_t, X_t)$

$$= \mathrm{Cov}(\phi X_{t-1} + Z_t,\ \phi X_{t-1} + Z_t)$$

$$= \phi^2 \gamma_X(0) + \sigma^2$$

Thus $\gamma_X(0) = \dfrac{\sigma^2}{1-\phi^2}$

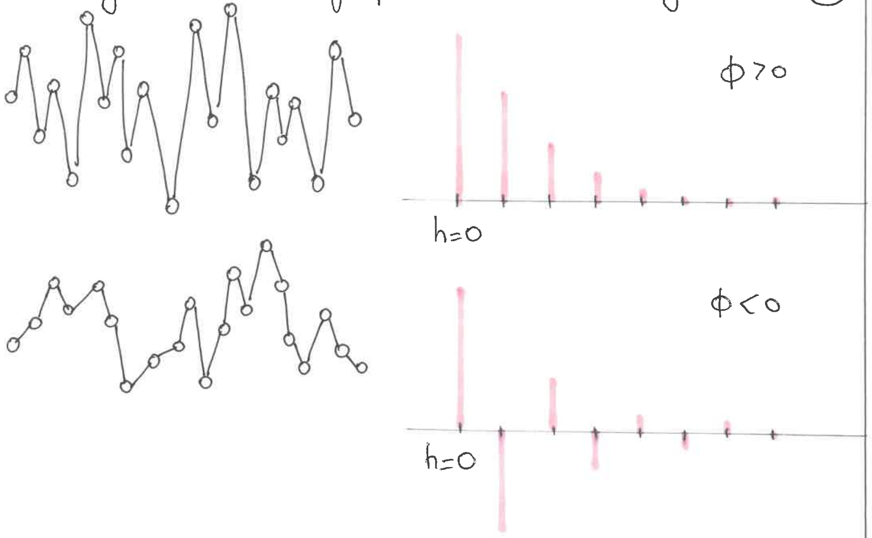
$$\boxed{\gamma_X(h) = \frac{\sigma^2}{1-\phi^2}\, \phi^{|h|} \quad \forall h}$$




dependence ↓ as the lag ↑

↳ the ACF of MA(1) and AR(1) processes exhibit very different behaviours.

When $\phi > 0$, value $X_t$ is similar to $X_{t-1}$ due to positive dependence → graph evolves smoothly. When $\phi < 0$, $\{X_t\}$ tends to switch signs, and the graph tends to be rougher. (7)




x <u>Remark</u>: ACF and LS prediction.

- Consider the best linear predictor of $X_{n+h}$ given $X_n$, when $\{X_n\}$ is stationary:

$$\boxed{\min_{\{\psi(X_n)=aX_n+b\}} \mathbb{E}\left(X_{n+h}-\psi(X_n)\right)^2}$$

$$= \text{a function of } a, b \equiv g(a,b).$$

Optimal values of $a$ and $b$ satisfy $\frac{\partial g(a,b)}{\partial a} = 0 = \frac{\partial g(a,b)}{\partial b}$.

After calculations, the best linear predictor is found to be $\psi(X_n) = \mathbb{E}X_n + \rho_X(h)(X_n - \mathbb{E}X_n)$, with MSE $\sigma^2(1-\rho_X(h)^2)$, where $\sigma^2 = \mathrm{Var}\, X_n$.

↑ The prediction accuracy improves as $|\rho_X(h)| \to 1$.

(8)

- Alternatively, if $(X_1, \dots, X_{n+h})$ is jointly Gaussian, the conditional distribution of $X_{n+h}$ given $X_n$ is $\mathcal{N}\left(\mathbb{E}X_{n+h} + \rho\frac{\sigma_{n+h}}{\sigma_n}(X_n - \mathbb{E}X_n),\ \sigma_{n+h}^2(1-\rho^2)\right)$.

So, for a Gaussian & stationary process $\{X_t\}$, the best predictor of $X_{n+h}$ given $X_n$ is

$$\mathbb{E}X_{n+h}|X_n = \operatorname*{argmin}_{\psi\in\mathcal{L}^2} \mathbb{E}\left(X_{n+h}-\psi(X_n)\right)^2$$

$$= \mathbb{E}X_n + \rho_X(h)(X_n - \mathbb{E}X_n),$$

and the MSE is $\sigma^2(1-\rho_X(h)^2)$.

⇒ If $\{X_t\}$ is stationary, $\psi(X_n) = \mathbb{E}X_n + \rho_X(h)(X_n - \mathbb{E}X_n)$ is the optimal linear predictor, and
If $\{X_t\}$ is stationary and Gaussian, $\psi$ is the optimal predictor.

↖ Linear prediction only needs second order statistics.

<u>Properties of ACF</u>:

(i) $\gamma(0) \geq 0$ ← Since $\mathrm{Var}\, X_t \geq 0$

(ii) $|\gamma(h)| \leq \gamma(0)$

(iii) $\gamma(h) = \gamma(-h)$

(iv) $\gamma$ is positive semi-definite

CS inequality
$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h})$
$\leq \sqrt{\mathrm{Var}\, X_t\ \mathrm{Var}\, X_{t+h}}$
$= \gamma(0)$

$\sum_{i,j}^{n} a_i \gamma(i-j) a_j \geq 0$
$\| \mathrm{Var}\ \underline{a}^t \underline{X}_n$

$\underline{a} = (a_1, \dots, a_n)^t$
$\underline{X}_n = (X_1, \dots, X_n)^t$

## I.2. Estimating the ACF.

- Having observed $x_1, \dots, x_n$, the sample mean is $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$, and the sample autocovariance function is

$$\hat{\gamma}(h) := \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}) \quad , \; -n < h < n$$

The sample autocorrelation function is $\hat{\rho}(h) = \dfrac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$.

- For any sequence $x_1, \dots, x_n$, the sample autocovariance function $\hat{\gamma}$ satisfies (i) $\hat{\gamma}(h) = \hat{\gamma}(-h)$
  (ii) $\hat{\gamma}$ is positive semi-definite
  (iii) $\hat{\gamma}(0) \geq 0$ & $|\hat{\gamma}(h)| \leq \hat{\gamma}(0)$.

Indeed, with $\tilde{X}_t = X_t - \mu_X$ and

$$\underset{n\times(2n-1)}{M} := \begin{pmatrix} 0 & \cdots & 0 & 0 & \tilde{X}_1 & \tilde{X}_2 & \cdots & \tilde{X}_{n-1} & \tilde{X}_n \\ 0 & \cdots & 0 & \tilde{X}_1 & \tilde{X}_2 & \tilde{X}_3 & \cdots & \tilde{X}_n & 0 \\ 0 & \cdots & \tilde{X}_1 & \tilde{X}_2 & \tilde{X}_3 & \tilde{X}_4 & \cdots & 0 & 0 \\ & & \vdots & \vdots & & \vdots & & & \\ \tilde{X}_1 & \cdots & \tilde{X}_{n-2} & \tilde{X}_{n-1} & \tilde{X}_n & 0 & \cdots & 0 & 0 \end{pmatrix}$$

we have that

$$\underline{\underline{\Gamma}}_n := \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(n-2) \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \cdots & \hat{\gamma}(0) \end{pmatrix} = \frac{1}{n} M M^t$$

$$\Rightarrow a^t \underline{\underline{\Gamma}}_n a = \frac{1}{n}\|M^t a\|^2 \geq 0$$



- For a stationary process $\{X_t\}$, the sample average $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ satisfies $\to E\,\bar{X}_n = \mu$

$$\to \operatorname{var} \bar{X}_n = \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h)$$

If $\sum_h |\gamma(h)| < \infty$, $\quad n \operatorname{var} \bar{X}_n \to \sum_{h\in\mathbb{Z}}\gamma(h) = \sigma^2 \underbrace{\sum_{h\in\mathbb{Z}}\rho(h)}_{\text{effect of correlation}}$

In the iid case, $\operatorname{var}\bar{X}_n \approx \dfrac{\sigma^2}{n}$.

In the correlated case, $\operatorname{var}\bar{X}_n \approx \dfrac{\sigma^2}{n/\tau}$ ; $\tau := \sum_{h\in\mathbb{Z}}\rho(h)$.

$\Rightarrow$ correlation $\equiv$ reduction of sample size from $n$ to $n/\tau$.

- Asymptotic results can be derived as well for the sample ACF, for the class of linear processes.
  A general framework for studying stationary ARMA processes (see later)

Def = A process is called a LINEAR PROCESS if it has the representation $X_t = \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j} \quad \forall t$, where $Z_t \sim WN(0, \sigma^2)$ and $\{\psi_t\}$ are constants such that $\sum_{j\in\mathbb{Z}} |\psi_j| < +\infty$

In fact, every second-order stationary process is either a linear process, or can be transformed to a linear process by substracting a deterministic component $\to$ WOLD DECOMPOS.

× **Remarks** : (i) The condition $\sum |\psi_j| < \infty$ ensures that the sum $\sum_{j\in\mathbb{Z}} \psi_j Z_{t-j}$ converges with probability 1 since



$$\mathbb{E}|X_t| \leq \sum_{j\in\mathbb{Z}} |\psi_j| \, \mathbb{E}|Z_{t-j}| \leq \sigma \sum_{j\in\mathbb{Z}} |\psi_j| < \infty$$

since $\mathbb{E} Z_t \leq \sigma \ \forall t$ (why ?)

and $\mathbb{E}|X_t| < \infty \Rightarrow |X_t| < \infty$ w.p. 1

(ii) In addition, this condition ensures that $\sum \psi_j^2 < \infty$, which in turn implies that the series converges in mean square ; i.e. $X_t$ is the MS limit of $\sum_{j=-n}^{n} \psi_j Z_{t-j}$, as $n \to +\infty$.

Indeed, a sequence $S_n$ of RVs converges in MS to some RV iff $\mathbb{E}(S_n - S_m)^2 \to 0$ as $m,n \to \infty$ [ in the Hilbert space of square integrable RVs ]

With $S_n = \sum_{j=-n}^{n} \psi_j Z_j$ , $\mathbb{E}(S_n - S_m)^2 = \mathbb{E}\Big(\sum_{m<|j|\leq n} \psi_j Z_j\Big)^2$,

$n > m > 0$

$$\mathbb{E}(S_n - S_m)^2 = \sigma^2 \sum_{m<|j|\leq n} \psi_j^2 \to 0 \text{ iff } \sum_{m<|j|\leq n} \psi_j^2 \to 0$$

as $m,n \to +\infty$

$\Updownarrow$

convergence of the sequence $\sum_{j=-n}^{n} \psi_j^2$

$\Leftrightarrow \sum_{j\in\mathbb{Z}} \psi_j^2 < +\infty$

For a linear process $X_t = \sum_j \psi_j Z_{t-j}$ , the ACF is $\gamma_X(h) = \sigma^2 \sum_j \psi_j \psi_{j+h}$



Since $\mathbb{E} X_t = 0$, and $\gamma_X(h) = \mathbb{E} X_t X_{t+h}$

$$= \mathbb{E}\Big(\sum_j \psi_j Z_{t-j}\Big)\Big(\sum_k \psi_k Z_{t+h-k}\Big)$$

$$= \sum_{j,k} \psi_j \psi_k \, \mathbb{E}(Z_{t-j} Z_{t+h-k})$$

$= 0$ unless $k = j-h$

$$= \sum_{j\in\mathbb{Z}} \psi_j \psi_{j+h} \, \sigma^2$$

**Theorem** : For a linear process $X_t = \sum_j \psi_j Z_{t-j}$ , $\sum |\psi_j| < \infty$,

(i) $n^{1/2} \bar{X}_n \xrightarrow{d} \mathcal{N}\Big(0, \sum_{h\in\mathbb{Z}} \gamma(h)\Big)$ ; $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

(ii) Under some additional moment conditions on the noise seq,

$$n^{1/2}\left\{ \begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(k) \end{pmatrix} - \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(k) \end{pmatrix} \right\} \xrightarrow{d} \mathcal{N}(0, \underline{V}),$$

where $\underline{V} = [V_{ij}]_{i,j}$,

$$V_{ij} = \sum_{h=1}^{\infty} [\rho(h+i) + \rho(h-i) - 2\rho(i)\rho(h)] \times [\rho(h+j) + \rho(h-j) - 2\rho(j)\rho(h)].$$

If $\rho(i) = 0$, $\forall i \neq 0$ then $\underline{V} = \underline{I}$

BARTLETT'S FORMULA

proof: (i) Recall from page 10 that $\operatorname{var}\overline{X}_n = \frac{1}{n}\sum\left(1-\frac{|h|}{n}\right)\gamma(h)$ 

$$\lim_{n\to\infty} n \operatorname{var}\overline{X}_n = \lim_{n\to\infty} \sum_{h=-(n-1)}^{n-1}\left(1-\frac{|h|}{n}\right)\gamma(h)$$

page 12

$$= \lim_{n\to\infty} \sigma^2 \sum_{j\in\mathbb{Z}} \psi_j \cdot \sum_{h=-(n-1)}^{n-1}\left(\psi_{j+h} - \frac{|h|}{n}\psi_{j+h}\right)$$

$$= \sigma^2\left(\sum_{j\in\mathbb{Z}}\psi_j\right)^2$$

$$\left(= \sigma^2 \sum_{h\in\mathbb{Z}} \gamma(h) \text{ from page 10}\right).$$

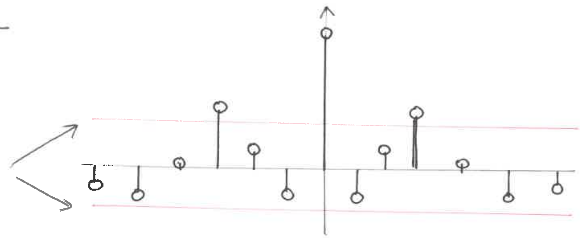
(ii) omitted ∎

- Consequence: Allows us to compute confidence intervals for $\hat{\gamma}(h)$.

If $\{X_t\}$ is a white noise sequence, we expect no more than $\approx 5\%$ of the peaks of the sample ACF to satisfy

$$|\hat{\rho}(h)| > \frac{1.96}{\sqrt{n}}$$

confidence bounds



As we shall see shortly, ARMA processes are linear processes, and we will be able to derive bounds on the sample ACF using the theorem of page 12.

↳ Useful to decide if $\hat{\gamma}(h)$ statistically deviates from zero.

# II. ARMA PROCESSES 

## II.1. AR(1) and causality.

Recall the definition of an AR(1) process, already encountered on page 6: $X_t = \phi X_{t-1} + Z_t$, $Z_t \sim WN(0,\sigma^2)$.

We assumed there that $|\phi| < 1$.

↳ It turns out that if $|\phi|<1$, $\boxed{X_t = \sum_{j=0}^{+\infty}\phi^j Z_{t-j}}$ (∗) (a linear process) is the unique stationary solution of $X_t = \phi X_{t-1} + Z_t$.

It is easy to see that (∗) satisfies ↑ since

$$\phi X_{t-1} = \sum_{j=0}^{+\infty}\phi^{j+1} Z_{t-1-j} = \sum_{j=1}^{+\infty}\phi^j Z_{t-j} \text{ so that}$$

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

To see that (∗) is the only stationary solution, let $\{Y_t\}$ be any solution (stat). Integrating (∗), we obtain

$$\begin{aligned} Y_t &= \phi Y_{t-1} + Z_t \\ &= Z_t + \phi Z_{t-1} + \phi^2 Y_{t-2} \\ &\;\vdots \\ &= Z_t + \phi Z_{t-1} + \cdots + \phi^k Z_{t-k} + \phi^{k+1} Y_{t-k-1} \end{aligned}$$

$$\Rightarrow Y_t - \sum_{j=0}^{k}\phi^j Z_{t-j} = \phi^{k+1} Y_{t-k-1}$$

$$E\left(\text{—\,''\,—}\right)^2 = \phi^{2(k+1)} \underbrace{E\, Y_{t-k-1}^2}_{\text{independent of } t} \to 0 \text{ as } k\to\infty$$

$\Rightarrow Y_t = $ MS limit of $\sum_{j=0}^{k}\phi^j Z_{t-j} \Rightarrow$ (∗) is the unique stationary solution. ∎

↳ In the case $|\phi| > 1$, the series $\sum_{j=0}^{+\infty} \phi^j Z_{t-j}$ does not converge, however, we can write:

$$X_t = -\phi^{-1} Z_{t+1} + \phi^{-1} X_{t+1}$$

$$= -\phi^{-1} Z_{t+1} - \phi^{-2} Z_{t+2} + \phi^{-2} X_{t+2}$$

$$\dots$$

$$= -\phi^{-1} Z_{t+1} - \dots - \phi^{-k-1} Z_{t+k+1} + \phi^{-k-1} X_{t+k+1}$$

(**) a linear process

& $\boxed{X_t = -\sum_{j=1}^{+\infty} \phi^{-j} Z_{t+j}}$ is the unique stationary solution of $X_t = \phi X_{t-1} + Z_t$, $|\phi| > 1$

unnatural, since $X_t$ is expressed in terms of the future values $Z_{t+1}, Z_{t+2}, \dots$ Solution (**) is said to be non-causal, as opposed to (*) in the case $|\phi| < 1$.

It is therefore customary to restrict attention to AR(1) process for which $|\phi| < 1$.

**Def** A linear process $\{X_t\}$ is CAUSAL if there is a function $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$ with $\sum |\psi_j| < \infty$, such that $X_t = \psi(B) Z_t$.

x Remark: When $|\phi| = 1$, there are no stationary solution to $X_t = \phi X_{t-1} + Z_t$.

Indeed, $X_t = X_{t-1} + Z_t$

$$= Z_t + \dots + Z_{t-k} + X_{t-k+1}$$

$$\Rightarrow X_t - X_{t-k-1} = \sum_{j=0}^{k} Z_{t-j}$$



Taking var(...) on both sides, (if stationary)

- $\operatorname{var}(X_t - X_{t-k-1}) = 2\gamma(0) - 2\gamma(k+1)$
- $\operatorname{var}\left(\sum_{j=0}^{k} Z_{t-j}\right) = \sum_{j=0}^{k} \operatorname{var} Z_{t-j} = (k+1)\sigma^2$.

$$\Rightarrow (k+1)\sigma^2 \leq 2\gamma(0) + 2\gamma(k+1) \leq 4\gamma(0)$$

$\Rightarrow \gamma(0) = +\infty$, a contradiction $\Rightarrow$ there are no stat. sol. ∎

**Summary**: A stationary solution of the AR(1) equations $X_t - \phi X_{t-1} = Z_t$, $Z_t \sim WN(0, \sigma^2)$ exists iff $|\phi| \neq 1$.

↘ If $|\phi| < 1$, the unique stationary solution is causal, and given by $X_t = \sum_{j=0}^{+\infty} \phi^j Z_{t-j}$.

↘ If $|\phi| > 1$, the unique stationary solution is noncausal, and given by $X_t = -\sum_{j=1}^{+\infty} \phi^{-j} Z_{t+j}$

## II.2. MA(1) and invertibility.

We already encountered the MA(1) process on page 5. It is defined as $X_t = Z_t + \theta Z_{t-1}$, $Z_t \sim WN(0, \sigma^2)$

$$= (1 + \theta B) Z_t$$

If $|\theta| < 1$, we can write

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

$$= (1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \cdots) X_t$$

$$= \sum_{j=0}^{+\infty} (-\theta)^j X_{t-j} = \text{a "causal" function of } X_t$$

We say that MA(1) is INVERTIBLE

If $|\theta|>1$, the sum $\sum_{j\geq 0}(-\theta)^j X_{t-j}$ diverges,
but we can write $Z_{t-1} = -\theta^{-1} Z_t + \theta^{-1} X_t$.

Just like the noncausal AR(1) process, we can show that

$$Z_t = -\sum_{j\geq 1}(-\theta)^{-j} X_{t+j} = \text{a linear function of } X_t.$$

↖ depends on the future values $X_{t+1}, X_{t+2}, \dots$

We say that MA(1) is <u>non-invertible</u>.

> <u>Summary</u> The stationary MA(1) process $X_t = Z_t + \theta Z_{t-1}$ is <u>causal</u>, and
> → <u>invertible</u> if $|\theta|<1$
> → <u>non-invertible</u> if $|\theta|>1$

## II.3. ARMA(1,1), causality & invertibility

The ARMA(1,1) process is defined as

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

$$\Leftrightarrow \underbrace{(1-\phi B)}_{=\phi(B)} X_t = \underbrace{(1+\theta B)}_{=\theta(B)} Z_t$$

<u>Q</u>: For which values of $\theta, \phi$ a stationary solution exists?

- If $|\phi|<1$, let $\chi(z)$ denote the power series expansion of $\frac{1}{\phi(z)} = \frac{1}{1-\phi z} = \sum_{j\geq 0}\phi^j z^j$ (absolutely summable).

$$\Rightarrow X_t = \underbrace{\chi(B)\theta(B)}_{=\psi(B)} Z_t \; ; \quad \psi(B) = (1+\phi B + \phi^2 B^2 + \dots)(1+\theta B) = \sum_{j\geq 0}\psi_j B^j$$



We see that $\begin{cases} \psi_0 = 1 \\ \psi_j = (\phi+\theta)\phi^{j-1}, & j\geq 1 \end{cases}$

$\Rightarrow$ The unique stationary solution is $X_t = Z_t + (\phi+\theta)\sum_{j\geq 1}\phi^{j-1} Z_{t-j}$ (causal) (linear process)

- Now, suppose $|\phi|>1$. We can represent $\frac{1}{\phi(z)}$ as a power series expansion in $z^{-1}$:

$$\frac{1}{\phi(z)} = -\sum_{j\geq 1}\phi^{-j} z^{-j}.$$

Using the same argument as before, we obtain the unique stationary solution

$$X_t = -\theta\phi^{-1} Z_t - (\theta+\phi)\sum_{j\geq 1}\phi^{-j-1} Z_{t+j}. \quad \text{(noncausal) (linear process)}$$

- If $|\phi| = \pm 1$, there is no stationary solution.
- A similar discussion applies to the notion of invertibility.

> <u>Summary</u>: A stationary solution of the ARMA(1,1) eq. $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$, $Z_t \sim WN(0,\sigma^2)$ exists iff $|\phi| \neq 1$.
> ↘ If $|\phi|<1$, the unique stat. solution is <u>causal</u>, and given by $X_t = Z_t + (\phi+\theta)\sum_{j\geq 1}\phi^{j-1} Z_{t-j}$
> ↘ If $|\phi|>1$, the unique stat. solution is <u>noncausal</u>, and given by $X_t = -\theta\phi^{-1} Z_t - (\theta+\phi)\sum_{j\geq 1}\phi^{-j-1} Z_{t+j}$
> ↘ If $|\theta|<1$, the ARMA(1,1) process is <u>invertible</u>.
> ↘ If $|\theta|>1$, the ARMA(1,1) process is <u>noninvertible</u>.

## II.4. AR(p), MA(q) & ARMA(p,q)



→ An AR(p) process $\{X_t\}$ is a stationary process that satisfies $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t$, $Z_t \sim WN(0,\sigma^2)$

$\Leftrightarrow \phi(B) X_t = Z_t$ where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$

↳ For which values of $\phi_1, \ldots, \phi_p$ a stationary solution exists?

- If $p=1$, we saw that a stationary solution exists iff $|\phi_1| \neq 1$. This is equivalent to the following condition on $\phi(z) = 1 - \phi_1 z$:

  $\forall z \in \mathbb{C}$, $\phi(z) = 0 \Rightarrow |z| \neq 1$

  To get stationarity, we want the roots of $\phi(z)$ to avoid the unit circle $\mathcal{C} = \{z \in \mathbb{C} : |z| = 1\}$.

> **Theorem:** A unique stationary solution to $\phi(B) X_t = Z_t$ exists iff $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \Rightarrow |z| \neq 1$.
> The AR(p) process is causal iff $\phi(z) = 0 \Rightarrow |z| > 1$

Roots outside the unit circle.

compare with $|\phi| < 1$ for AR(1)

→ The MA(q) process is defined as $X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$, $Z_t \sim WN(0, \sigma^2)$

It is possible to show that the stationary MA(q) process is invertible iff $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \neq 0 \ \forall |z| \leq 1$.

→ An ARMA(p,q) process is a stationary process that satisfies $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$, $Z_t \sim WN(0, \sigma^2)$.



Usually, we impose that $\phi_p, \theta_q \neq 0$, and that $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ & $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ have no common factors. This implies that it is not a lower order ARMA model.

> **Theorem:** If $\phi$ and $\theta$ have no common factors, a unique stationary solution to $\phi(B) X_t = \theta(B) Z_t$, $Z_t \sim WN(0,\sigma^2)$ exists iff the roots of $\phi(z)$ avoid the unit circle:
>
> $|z| = 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$.
>
> The ARMA(p,q) process is causal iff the roots of $\phi(z)$ are outside the unit circle:
>
> $|z| \leq 1 \Rightarrow \phi(z) \neq 0$
>
> It is invertible iff the roots of $\theta(z)$ are outside the unit circle:
>
> $|z| \leq 1 \Rightarrow \theta(z) \neq 0$

Moreover, it is possible to show that if $\forall |z| = 1$, $\theta(z) \neq 0$, then there exists $\tilde{\phi}$ and $\tilde{\theta}$ & $\tilde{W}_t \sim WN$ such that $\tilde{\phi}(B) X_t = \tilde{\theta}(B) Z_t$ which is causal & invertible.

⇒ We shall stick to causal & invertible ARMA processes.

## II.5. Autocovariance of ARMA processes.

We already calculated on pages 5 and 6 the ACF of an MA(1) and AR(1) processes. We discuss here two techniques for computing the ACF of a general ARMA(p,q).

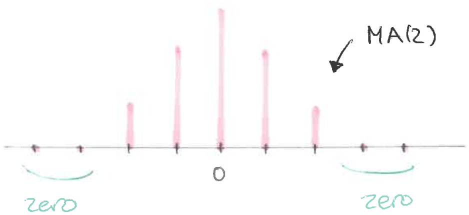
Ex: MA(q) process $X_t = \theta(B) Z_t$, where $Z_t \sim WN(0,\sigma^2)$ & $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. We readily get from the calculations on top of page 12 that

$$\boxed{\gamma_X(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|} & \text{if } |h| \le q \\ 0 & \text{if } |h| > q \end{cases}}$$

ACF of MA(q)

Vanishes for $h > q$.

Characteristic of a moving average process. In fact, it is possible to show that every zero-mean stationary process with correlations vanishing at lags $> q$ can be represented as a MA process of order q or less.




↳ Expression $\gamma(h) = \sigma^2 \sum_j \psi_j \psi_{j+h}$ for a linear process $X_t = \sum_j \psi_j Z_{t-j}$ can be used to derive the ACF of an ARMA process, once the coefficients $\psi_j$ are known. This can be tedious in the general case. As an alternative approach, notice that

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

$$\mathbb{E}\{(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}) X_{t-h}\}$$

$$= \mathbb{E}\{(Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}) X_{t-h}\}$$



$$\Leftrightarrow \gamma(h) - \phi_1 \gamma(h-1) - \cdots - \phi_p \gamma(h-p)$$

$$= \mathbb{E}(\theta_h Z_{t-h} X_{t-h} + \cdots + \theta_q Z_{t-q} X_{t-h})$$

$$= \sigma^2 \sum_{j=0}^{q-h} \theta_{h+j} \psi_j \quad (\text{putting } \theta_0 = 1)$$

The RHS vanishes for most values of h

↳ Homogeneous linear difference equations with constant coefficients.

Ex: Consider $(1 + \frac{1}{4} B^2) X_t = (1 + \frac{1}{5} B) Z_t$.

$$\Leftrightarrow X_t = \psi(B) Z_t \text{ with } \psi_j = \left(1, \tfrac{1}{5}, -\tfrac{1}{4}, -\tfrac{1}{20}, \dots\right)$$

Then

$$\gamma(h) - \phi_1 \gamma(h-1) - \phi_2 \gamma(h-2) = \sigma^2 \sum_{j=0}^{q-h} \theta_{h+j} \psi_j$$

$$\Updownarrow$$

$$\gamma(h) + \frac{1}{4}\gamma(h-2) = \begin{cases} \sigma^2(\psi_0 + \frac{1}{5}\psi_1) & \text{if } h = 0 \\ \frac{1}{5}\sigma^2 \psi_0 & \text{if } h = 1 \\ 0 & \text{o/w} \end{cases}$$

We need to solve $\gamma(h) + \frac{1}{4}\gamma(h-2) = 0 \quad \forall h \geq 2$, with initial conditions

$$\begin{cases} \gamma(0) + \frac{1}{4}\gamma(-2) = \sigma^2(1 + 1/25) \\ \gamma(1) + \frac{1}{4}\gamma(-1) = \sigma^2/5 \end{cases}$$

× Linear Difference Equations: $a_0 x_t + a_1 x_{t-1} + \cdots + a_k x_{t-k} = 0$

$$\Leftrightarrow a(B) x_t = 0$$

with $a(B) = a_0 + a_1 B + \cdots + a_k B^k$.

Consider the auxiliary equation (aka characteristic polynomial)

$a(z) = a_0 + a_1 z + \cdots + a_k z^k$, $z \in \mathbb{C}$

$= (z - z_1)(z - z_2) \times \cdots \times (z - z_k)$, $z_1, \dots, z_k \in \mathbb{C}$.

There are three cases:

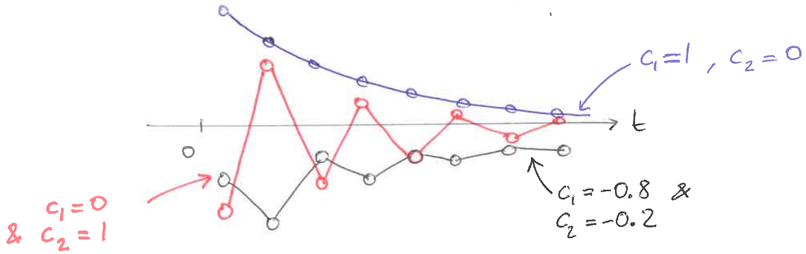
- The $z_j$ are real and distinct.

Then $x_t = c_1 z_1^{-t} + \dots + c_k z_k^{-t}$

$= $ linear combination of solutions to $(B - z_j) x_t = 0$

Describes already a wide range of possible dynamics.

Ex: $k = 2$, $z_1 = 1{,}2$ & $z_2 = -1.3$




- The $z_j$ are complex and distinct

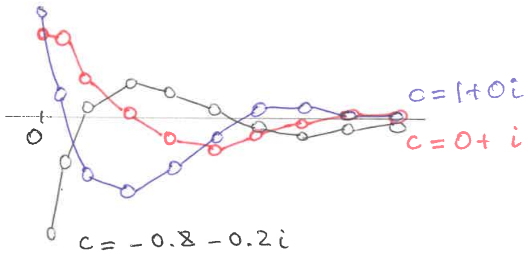
Since $a_i \in \mathbb{R}$, if $z_j$ is solution, then so is the complex conjugate $\bar{z}_j$. For $x_t$ to be real, the coefficient in front of $\bar{z}_j^{-t}$ must be $\bar{c}_j$.

$$\hookrightarrow x_t = c z_1^{-t} + \bar{c} \bar{z}_1^{-t}$$

$$= r e^{i\theta} |z_1|^{-t} e^{-i\omega t} + r e^{-i\theta} |z_1|^{-t} e^{i\omega t}$$

$$= 2r |z_1|^{-t} \cos(\omega t - \theta)$$

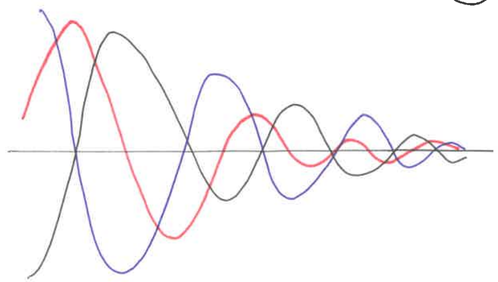
Ex: $z_1 = 1.2 + i$, $\bar{z}_1 = 1.2 - i = z_2$






Ex: $z_1 = 1 + 0.1 i = \bar{z}_2$

oscillating behaviour



- Some $z_j$ are repeated.

Can show that $(B - z_1)^m x_t = 0$ has solution $(c_0 + c_1 t + \dots + c_{m-1} t^{m-1}) z_1^{-t}$

More generally,

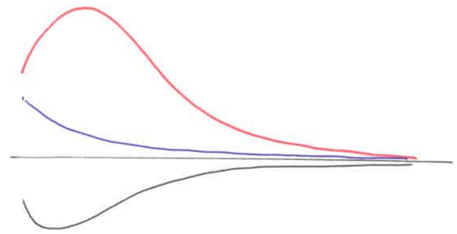
$(z - z_1)^{m_1} \dots (z - z_\ell)^{m_\ell} = 0$

has solution

$c_1(t) z_1^{-t} + \dots + c_\ell(t) z_\ell^{-t}$

where

$c_i(t)$ is a polynomial of degree $m_i - 1$.



Back to the example on page 22, the characteristic polynomial $1 + \frac{1}{4} z^2 = \frac{1}{4}(z - 2i)(z + 2i)$ has roots $z_1 = 2 e^{i\pi/2} = \bar{z}_2$.

The solution has the form $\gamma(h) = c z_1^{-h} + \bar{c} \bar{z}_1^{-h}$

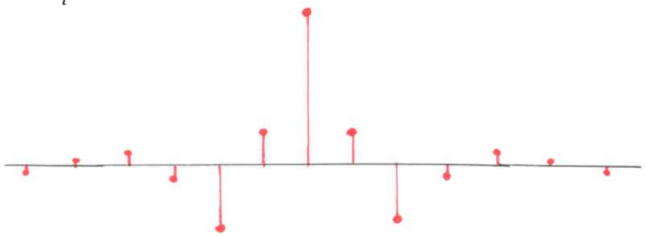
$$= 2^{-h} \left( |c| e^{i(\theta - \frac{h\pi}{2})} + |c| e^{i(-\theta + \frac{h\pi}{2})} \right)$$

$$= c_1 2^{-h} \cos\left( \frac{h\pi}{2} - \theta \right)$$

where $c_1$ & $\theta$ are determined from the initial conditions

$\gamma(0) + \frac{1}{4}\gamma(2) = \sigma^2 \left(1 + \frac{1}{25}\right)$ and $1.25\, \gamma(1) = \frac{\sigma^2}{5}$.

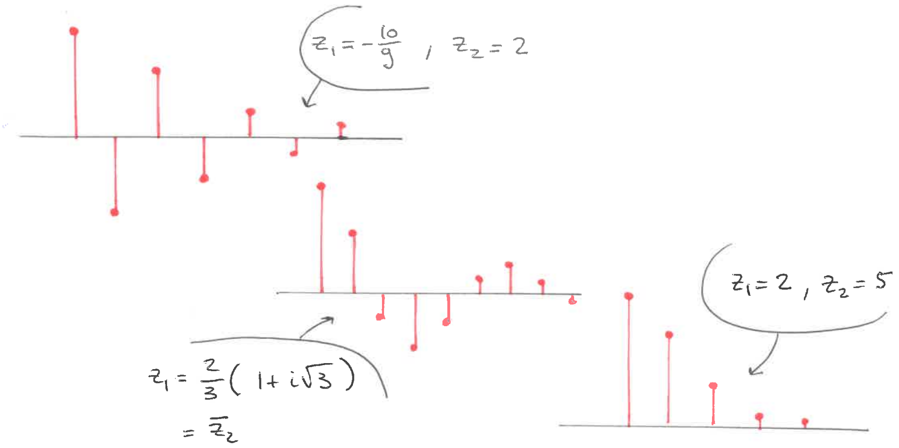
$\Rightarrow$ Plug $\gamma(0) = c_1 \cos\theta$, $\gamma(1) = \frac{c_1}{2}\sin\theta$ and $\gamma(2) = -\frac{c_1}{4}\cos\theta$ into the initial conditions. (25)



The roots of $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ determine
- → if the process is stationary (no roots on $\mathcal{C}$)
- → if the process is causal (roots outside $\mathcal{C}$)
- → the shape of the ACF (oscillatory, damped, ...)

Ex: AR(2) process $(1 - z_1^{-1}B)(1 - z_2^{-1}B)X_t = Z_t$, $|z_1|, |z_2| > 1$ (causal & stationary)

Depending on the values of $z_1, z_2$, the ACF may look like:




Ex: ARMA(1,1). (26)

Recall from calculations pages 17/18 that a causal ARMA(1,1) process can be expressed as $X_t = Z_t + (\phi+\theta)\sum_{j\geq 1}\phi^{j-1}Z_{t-j}$, so that $\psi_0 = 1$ and $\psi_j = (\phi+\theta)\phi^{j-1}$, $j \geq 1$. We may use $\gamma(h) = \sigma^2 \sum_j \psi_j \psi_{j+h}$ directly here.

- $\gamma(0) = \sigma^2 \sum_j \psi_j^2 = \sigma^2\left(1 + (\phi+\theta)^2 \sum_{j\geq 0}\phi^{2j}\right)$

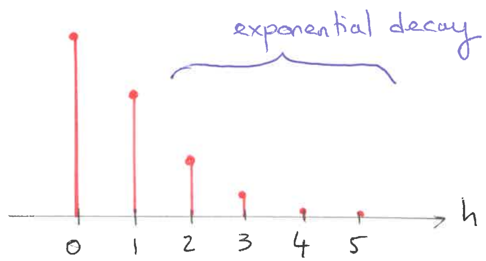
$$= \sigma^2\left(1 + \frac{(\phi+\theta)^2}{1-\phi^2}\right)$$

- $\gamma(1) = \sigma^2 \sum_{j\geq 0}\psi_j\psi_{j+1}$

$$= \sigma^2\Big(\underbrace{\theta+\phi}_{\psi_0\psi_1} + (\theta+\phi)^2\phi\sum_{j\geq 0}\phi^{2j}\Big)$$

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

- $\gamma(h) = \phi^{h-1}\gamma(1)$, $h \geq 2$




Compare with the expressions of the ACF of an MA(1) and AR(1).
The ACF of ARMA(1,1) ≈ mixture of them: expression $\gamma(0)$ and $\gamma(1)$ in terms of $\sigma^2/\phi/\theta$, followed by an exponential decay.

# III. FORECASTING STATIONARY PROCESSES

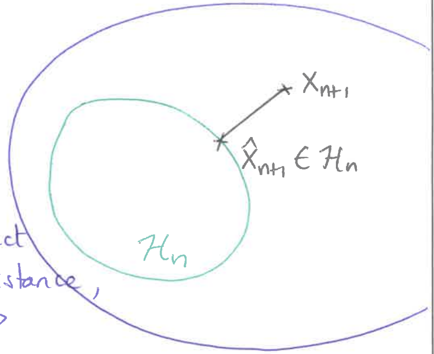
## III.1- step ahead prediction.

Let $X_1, \dots, X_n$ be a zero-mean stationary time-series with known autocovariance function $\gamma$. Having observed $X_1, \dots, X_n$, the goal is to predict $X_{n+1} \to$ denoted $\hat{X}_{n+1}$. We restrict ourselves to linear combinations of $X_1, \dots, X_n$.

Consider $\mathcal{H}_n := \overline{sp}\{X_1, \dots, X_n\}$
$=$ closed linear subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$

" Best " prediction $\equiv$ prediction minimizing the MSE.

$$
\begin{aligned}
MSE(Y) &= \mathbb{E}(X_{n+1} - Y)^2 \\
&= \| X_{n+1} - Y \|^2 \\
&= d^2(X_{n+1}, Y)
\end{aligned}
$$



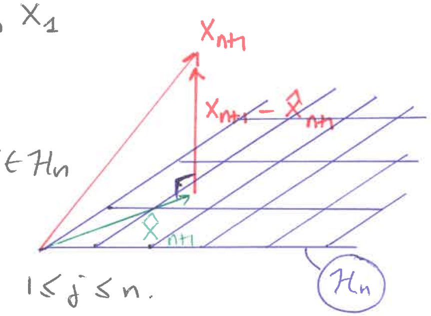

Since $\langle X, Y\rangle := \mathbb{E}XY$ = inner product
& $d^2(X,Y) = \|X - Y\|^2$ = (sq) distance,
where $\|X\|^2 = \mathbb{E}X^2 = \langle X, X\rangle$
in the space $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ of square integrable RVs.

↳ the projection theorem ensures the existence of a unique point $\hat{X}_{n+1}$ in $\mathcal{H}_n$ for which $\|X_{n+1} - Y\|$ is minimized over $\mathcal{H}_n$. ($Y \in \mathcal{H}_n$)

$\hat{X}_{n+1}$ is the orthogonal projection of $X_{n+1}$ onto $\mathcal{H}_n$.
We write $\hat{X}_{n+1} = P_{\mathcal{H}_n} X_{n+1}$.



Since $\hat{X}_{n+1} \in \mathcal{H}_n$, we can write
$\hat{X}_{n+1} = \phi_{n,1} X_n + \dots + \phi_{n,n} X_1$




Note that
$\langle X_{n+1} - \hat{X}_{n+1}, Y\rangle = 0 \quad \forall Y \in \mathcal{H}_n$
or, equivalently,
$\langle X_{n+1} - \hat{X}_{n+1}, X_j\rangle = 0, \quad 1 \le j \le n.$

$$\left\langle X_{n+1} - \sum_{i=1}^{n} \phi_{n,i} X_{n+1-i},\; X_{n+1-j} \right\rangle = 0, \quad 1 \le j \le n$$

$$\underbrace{\langle X_{n+1}, X_{n+1-j}\rangle}_{= \gamma(j)} = \sum_{i=1}^{n} \phi_{n,i} \underbrace{\langle X_{n+1-i}, X_{n+1-j}\rangle}_{= \gamma(i-j)}$$

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

$$
\underbrace{\begin{pmatrix}
\gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(n-1) \\
\gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(n-2) \\
\gamma(2) & \gamma(1) & \gamma(0) & \dots & \gamma(n-3) \\
\vdots & \vdots & \vdots & & \vdots \\
\gamma(n-1) & \gamma(n-2) & \gamma(n-3) & \dots & \gamma(0)
\end{pmatrix}}_{=: \underline{\Gamma}_n}
\underbrace{\begin{pmatrix}
\phi_{n,1} \\ \phi_{n,2} \\ \phi_{n,3} \\ \vdots \\ \phi_{n,n}
\end{pmatrix}}_{=: \underline{\phi}_n}
=
\underbrace{\begin{pmatrix}
\gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \vdots \\ \gamma(n)
\end{pmatrix}}_{=: \underline{\gamma}_n}
$$

(covariance matrix of $\underline{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$)

$$\boxed{\underline{\Gamma}_n \underline{\phi}_n = \underline{\gamma}_n}$$

The projection theorem guarantees that $\underline{\underline{\Gamma}}_n \underline{\phi}_n = \underline{\gamma}_n$ has at least one solution, since $\hat{X}_{n+1}$ must be expressible in terms of $X_1, \dots, X_n$. In fact, if $\gamma(0) > 0$ and $\gamma(h) \to 0$ as $h \to \infty$, then one can show that $\underline{\underline{\Gamma}}_n$ is non-singular $\forall n$, and $\underline{\phi}_n = \underline{\underline{\Gamma}}_n^{-1} \underline{\gamma}_n$ (see proposition 5.1.1. in [TSTM]).

requires an $(n \times n)$ matrix inversion.

## III. 2. h-step ahead prediction.

The best linear predictor of $X_{n+h}$ given $X_1, \dots, X_n$ can be found in exactly the same way.

$$\hat{X}_{n+h} = P_{\mathcal{H}_n} X_{n+h} =: \phi_{n,1}^{(h)} X_n + \dots + \phi_{n,n}^{(h)} X_1,$$

where the coefficients $\phi_{n,j}^{(h)}$, $j = 1, \dots, n$ satisfy

$$\underline{\underline{\Gamma}}_n \underline{\phi}_n^{(h)} = \underline{\gamma}_n^{(h)},$$

with

$$\underline{\gamma}_n^{(h)} = \begin{pmatrix} \gamma(h) \\ \vdots \\ \gamma(n+h-1) \end{pmatrix}, \quad \underline{\phi}_n^{(h)} = \begin{pmatrix} \phi_{n,1}^{(h)} \\ \vdots \\ \phi_{n,n}^{(h)} \end{pmatrix}$$

## III. 3. The Durbin-Levinson Algorithm.

The computation of $\hat{X}_{n+1}$ requires the inversion of an $(n \times n)$ matrix. The goal is to derive a recursive algorithm that (i) avoids matrix inversions and (ii) makes use of $\hat{X}_n = P_{\mathcal{H}_{n-1}} X_n$ to compute $\hat{X}_{n+1} = P_{\mathcal{H}_n} X_{n+1}$.

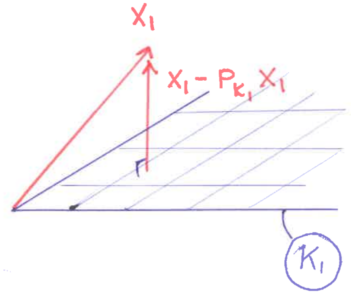
Put $v_n := \mathbb{E}(X_{n+1} - \hat{X}_{n+1})^2$



Note that $\hat{X}_{n+1} = P_{\mathcal{H}_n} X_{n+1} \in \mathcal{H}_n = \underbrace{\overline{sp}\{X_1, \dots, X_n\}}$.

The idea is to decompose $\mathcal{H}_n$ into two $\perp$ subspaces, one of them spanned by $(n-1)$ variables

basis of our recursion.




let

$K_1 := \overline{sp}\{X_2, \dots, X_n\}$

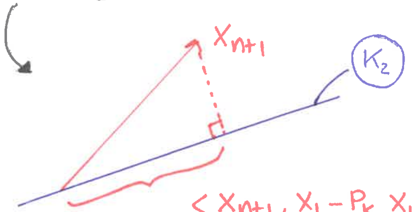
$K_2 := \overline{sp}\{X_1 - P_{K_1} X_1\}$

$\Rightarrow K_1 \perp K_2 \qquad K_1 \cap K_2 = \{0\}$

$\Rightarrow \hat{X}_{n+1}$ can be decomposed as a sum of two terms:

$$\hat{X}_{n+1} = P_{K_1} X_{n+1} + P_{K_2} X_{n+1}$$

$$= \underbrace{P_{K_1} X_{n+1}}_{\in K_1} + \underbrace{a(X_1 - P_{K_1} X_1)}_{\in K_2}$$




$$a = \frac{\langle X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\| X_1 - P_{K_1} X_1 \|^2}$$

$\Rightarrow$ We need to compute $P_{K_1} X_{n+1}$ and $P_{K_1} X_1$.

Recall that $\underline{\underline{\Gamma}}_{n-1}$ denotes the covariance matrix of $(X_1, \dots, X_{n-1})$.

By stationarity, the random vectors $(X_{n-1}, \dots, X_1)$ and $(X_2, \dots, X_n)$ also have cov. matrix $\underline{\underline{\Gamma}}_{n-1}$

reverse order

shift

We can thus write

$$\begin{cases} P_{K_1} X_1 = \phi_{n-1,1} X_2 + \cdots + \phi_{n-1,n-1} X_n = \sum_{j=1}^{n-1} \phi_{n-1,j} X_{j+1} \\ P_{K_1} X_{n+1} = \phi_{n-1,1} X_n + \cdots + \phi_{n-1,n-1} X_2 = \sum_{j=2}^{n-1} \phi_{n-1,j} X_{n+1-j} \end{cases}$$

instead of $X_{n-1}$ - - - $X_1$,
since they solve the same system of equations $\underline{\Gamma_n}\,\underline{\Phi_n} = \underline{\gamma_n}$.

$$\Rightarrow \hat{X}_{n+1} = a X_1 + \sum_{j=1}^{n-1} \left( \phi_{n-1,j} - a\,\phi_{n-1,n-j} \right) X_{n+1-j}, \quad (*)$$

where

$$a = \left( \langle X_{n+1}, X_1 \rangle - \sum_{j=1}^{n-1} \phi_{n-1,j} \langle X_{n+1}, X_{j+1} \rangle \right) \| X_1 - P_{K_1} X_1 \|^{-2}.$$

Again, from stationarity, predicting $X_1$ from $X_2, \dots, X_n$ yields the same error as predicting $X_n$ from $X_1, \dots, X_{n-1}$ (since the two problems yield the same coefficients $\phi_{n-1,1}, \dots, \phi_{n-1,n-1}$).

Thus,

$$\| X_1 - P_{K_1} X_1 \|^2 = \| X_n - P_{\mathcal{H}_{n-1}} X_n \|^2 = v_{n-1}, \text{ by definition.}$$

We obtain

$$a = \left\{ \gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\, \gamma(n-j) \right\} v_{n-1}^{-1}.$$

Equating coefficients in (*) with $\hat{X}_{n+1} = \phi_{n,1} X_n + \cdots + \phi_{n,n} X_1$, we see that

$$\boxed{\begin{aligned} &\phi_{n,n} = a = \left\{ \gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\, \gamma(n-j) \right\} v_{n-1}^{-1} \\ \& \quad &\phi_{n,j} = \phi_{n-1,j} - \phi_{n,n}\, \phi_{n-1,n-j}, \quad 1 \le j \le n-1 \end{aligned}}$$



It remains to derive a recursion for $v_n$.

$$\begin{aligned} v_n &= \| X_{n+1} - \hat{X}_{n+1} \|^2 \\ &= \| \underbrace{X_{n+1} - P_{K_1} X_{n+1}} - \underbrace{P_{K_2} X_{n+1}} \|^2 \\ &= \| X_{n+1} - P_{K_1} X_{n+1} \|^2 + \| P_{K_2} X_{n+1} \|^2 - 2 \langle X_{n+1} - P_{K_1} X_{n+1}, P_{K_2} X_{n+1} \rangle \end{aligned}$$

$\perp$

$P_{K_2} X_{n+1} = a(X_1 - P_{K_1} X_1)$

$$= v_{n-1} + a^2 v_{n-1} - 2a \langle X_{n+1}, X_1 - P_{K_1} X_1 \rangle$$

since $P_{K_2} X_{n+1} = a(X_1 - P_{K_1} X_1) \Rightarrow \| P_{K_2} X_{n+1} \|^2 = a^2 \| X_1 - P_{K_1} X_1 \|^2 = a^2 v_{n-1}$

Since by definition $a = \dfrac{\langle X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\| X_1 - P_{K_1} X_1 \|^2}$, we get

$$\underline{v_n} = v_{n-1} + a^2 v_{n-1} - 2a^2 v_{n-1} = \underline{v_{n-1}(1 - a^2)} \quad \blacksquare$$

## SUMMARY: DURBIN-LEVINSON ALGORITHM.

$\{X_t\}$ = zero-mean stationary process with ACF $\gamma$, $\gamma(0) > 0$

Then $\hat{X}_{n+1} = \phi_{n,1} X_n + \cdots + \phi_{n,n} X_1$, with

$$\begin{cases} \phi_{n,n} = \left\{ \gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\, \gamma(n-j) \right\} v_{n-1}^{-1} \\ \phi_{n,j} = \phi_{n-1,j} - \phi_{n,n}\, \phi_{n-1,n-j}, \quad 1 \le j \le n-1 \\ v_n = v_{n-1} (1 - \phi_{n,n}^2) \end{cases}$$

MSE

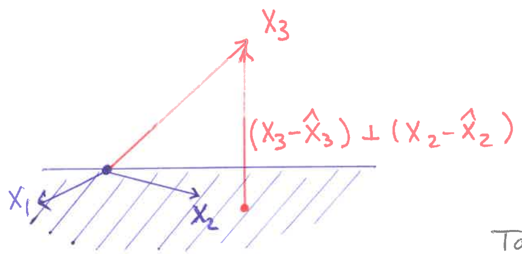
+ initial cond. $\phi_{1,1} = \dfrac{\gamma(1)}{\gamma(0)}$, $v_0 = \gamma(0)$.

## III.4. The innovations algorithm.

Let $\{X_t\}$ be a possibly non-stationary, zero-mean process, with covariance function $\kappa(i,j) := \mathbb{E}\, X_i X_j$. The central idea here is to decompose $\hat X_{n+1}$ in terms of "innovations" $(X_1-\hat X_1),\dots,(X_n-\hat X_n)$, instead of $X_1,\dots,X_n$:

$$\hat X_{n+1} = \sum_{j=1}^{n} \theta_{n,j}\,(X_{n+1-j}-\hat X_{n+1-j}) \quad \text{———} (*)$$

since $\mathcal{H}_n = \overline{sp}\{X_1,\dots,X_n\} = \overline{sp}\{X_1-\hat X_1,\dots,X_n-\hat X_n\}$




$(X_i-\hat X_i)\in\mathcal{H}_{j-1}$, $i<j$

&

$(X_j-\hat X_j)\perp\mathcal{H}_{j-1}$ by def

Take the inner product on both sides of $(*)$ with $X_{k+1}-\hat X_{k+1}$, $0\le k\le n-1$.

$$\langle \hat X_{n+1}, X_{k+1}-\hat X_{k+1}\rangle = \sum_{j=1}^{n}\theta_{n,j}\langle X_{n+1-j}-\hat X_{n+1-j}, X_{k+1}-\hat X_{k+1}\rangle$$

$\perp$ unless $n+1-j=k+1$, i.e. $j=n-k$

$$\langle \hat X_{n+1}, X_{k+1}-\hat X_{k+1}\rangle = \theta_{n,n-k}\,\|X_{k+1}-\hat X_{k+1}\|^2.$$

$$\theta_{n,n-k} = v_k^{-1}\langle X_{n+1}, X_{k+1}-\hat X_{k+1}\rangle$$

Since $X_{n+1}-\hat X_{n+1}\perp X_{k+1}-\hat X_{k+1}$



We get

$$\theta_{n,n-k} = v_k^{-1}\Big(\underbrace{\langle X_{n+1}, X_{k+1}\rangle}_{=\,\kappa(n+1,k+1)} - \langle X_{n+1}, \underbrace{\hat X_{k+1}}_{=\,\sum_{j=0}^{k-1}\theta_{k,k-j}(X_{j+1}-\hat X_{j+1})}\rangle\Big) \qquad (k=0,\dots,n-1)$$

$$\theta_{n,n-k} = v_k^{-1}\Big(\kappa(n+1,k+1) - \sum_{j=0}^{k-1}\theta_{k,k-j}\underbrace{\langle X_{n+1}, X_{j+1}-\hat X_{j+1}\rangle}\Big)$$

Also, we have $v_j\,\theta_{n,n-j} = \langle X_{n+1}, X_{j+1}-\hat X_{j+1}\rangle$.

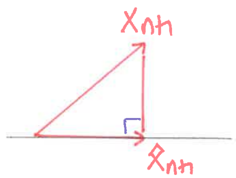
Thus

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

Compute $\theta_{n,n}\to\theta_{n,n-1}\to\theta_{n,n-2}\to\cdots\to\theta_{n,1}$

In addition, $v_0 = \|X_1-\hat X_1\|^2 \underset{\hat X_1=0}{=} \mathbb{E}\,X_1^2 = \kappa(1,1)$

& $v_n = \|X_{n+1}-\hat X_{n+1}\|^2$
$= \|X_{n+1}\|^2 - \|\hat X_{n+1}\|^2$




Since

$$\hat X_{n+1} = \sum_{k=0}^{n-1}\theta_{n,n-k}\,(X_{k+1}-\hat X_{k+1})$$

(set $j=n-k$ in $(*)$ page 33)

$\Downarrow$

$\|\hat X_{n+1}\|^2 = \sum_{k=0}^{n-1}\theta_{n,n-k}^2\,v_k$, we finally obtain:

$$\boxed{v_n = \kappa(n+1,n+1) - \sum_{k=0}^{n-1}\theta_{n,n-k}^2\,v_k}$$

$v_0\to v_1\to v_2\cdots$

## SUMMARY: THE INNOVATIONS ALGORITHM

$\{X_t\}$ = zero-mean process with $\kappa(i,j) = \mathbb{E}(X_i X_j)$.

Then $\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j})$, $n \geq 1$, $\hat{X}_1 = 0$,

with

$$\begin{cases} v_0 = \kappa(1,1) \\ \theta_{n,n-k} = v_k^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\,\theta_{n,n-j}\,v_j\right) & n \geq 1,\ k = 0,\dots,n-1 \\ v_n = \kappa(n+1,n+1) - \sum_{k=0}^{n-1} \theta_{n,n-k}^2\, v_k \end{cases}$$

$v_0 \to \theta_{11} \to v_1 \to \theta_{22} \to \theta_{21} \to v_2 \to \theta_{33} \to \theta_{32} \to \theta_{31} \to \dots$

× Remark: Recursive calculation for $h$-step prediction, $h \geq 1$.

$$P_{\mathcal{H}_n} X_{n+h} = P_{\mathcal{H}_n}\Big(\underbrace{P_{\mathcal{H}_{n+h-1}} X_{n+h}}_{=\ 1\text{-step ahead pred}}\Big)$$

$$= P_{\mathcal{H}_n} \hat{X}_{n+h}$$

$$= P_{\mathcal{H}_n}\left(\sum_{j=1}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j})\right)$$

$X_{n+h-j} - \hat{X}_{n+h-j} \perp \mathcal{H}_n$ for $j < h$.

$$\Rightarrow \boxed{P_{\mathcal{H}_n} X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j})}$$

with
$$\text{MSE} = \mathbb{E}(X_{n+h} - P_{\mathcal{H}_n} X_{n+h})^2$$
$$= \|X_{n+h}\|^2 - \|P_{\mathcal{H}_n} X_{n+h}\|^2$$
$$= \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2\, v_{n+h-j-1}.$$



## III.5. Forecasting an ARMA(p,q) process.

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

Then
$$\begin{cases} \kappa(i,j) = 0 & \text{for } |i-j| > 1 \\ \kappa(i,i) = \sigma^2(1+\theta^2) \\ \kappa(i,i+1) = \theta\sigma^2 \end{cases}$$
(see page 5)

The innovations algorithm simplifies to:

$$\begin{cases} v_0 = \sigma^2(1+\theta^2) \\ \theta_{n,j} = \begin{cases} v_{n-1}^{-1}\theta\sigma^2 & \text{if } j = 1 \\ 0 & \text{if } j = 2,\dots,n \end{cases} \\ v_n = (1 + \theta^2 - v_{n-1}^{-1}\theta^2\sigma^2)\,\sigma^2 \end{cases}$$

Thus $\hat{X}_{n+1} = \theta_{n,1}(X_n - \hat{X}_n)$

$$\hat{X}_{n+1} = v_{n-1}^{-1}\sigma^2\ \boxed{\theta\,(X_n - \hat{X}_n)}$$

$\approx \theta Z_{t-1}$ in $X_t = Z_t + \theta Z_{t-1}$

innovation $X_n - \hat{X}_n \approx$ noise term

Note that most of the coefficients $\theta_{n,j}$ vanish for an MA(1) process: only $\theta_{n,1} > 0$.

More generally, we conclude similarly that $\theta_{n,j} = 0$, $j > q$ for an MA(q) process.

↳ The innovations algorithm is convenient to predict MA(q) processes.

- Causal ARMA$(p,q)$ $\quad \phi(B)X_t = \theta(B)Z_t$, $\quad Z_t \sim WN(0, \sigma^2)$
  zero-mean

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}.$$

We apply the innovations algorithm to the transformed process

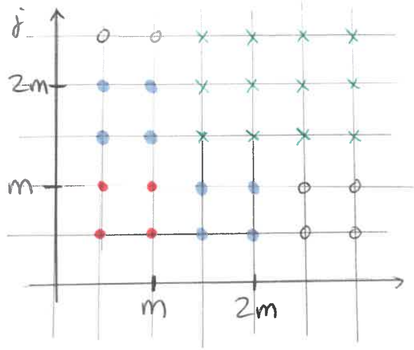
$$\begin{cases} W_t := \sigma^{-1} X_t & t = 1, \ldots, m := \max(p,q) \\ W_t := \sigma^{-1}\phi(B) X_t, & t > m \end{cases}$$

(non-stationary)

Put $\theta_0 := 1$ and $\theta_j := 0$ for $\forall j > q$.

Let $\gamma$ denote the ACF of $X_t$.

$\to$ We derive the expression of the ACF of $W_t$ in terms of $\gamma$.




$\leftarrow$ let $K(i,j) = \mathbb{E}\, W_i W_j$ $\quad \forall i,j \geq 1$.

Partition the space $\{(i,j),\ i,j \geq 1\}$ in 4 subspaces ( • • × ○ )

- $1 \leq i, j \leq m$ : $K(i,j) = \sigma^{-2}\mathbb{E}\, X_i X_j = \sigma^{-2}\gamma(i-j)$.

- $\min(i,j) \leq m < \max(i,j) \leq 2m$

  Take $i < j$

$$\begin{aligned} K(i,j) &= \mathbb{E}\, W_i W_j \\ &= \mathbb{E}\,(\sigma^{-1} X_i)(\sigma^{-1}\phi(B) X_j) \\ &= \sigma^{-2}\,\mathbb{E}\{X_i(X_j - \phi_1 X_{j-1} - \cdots - \phi_p X_{j-p})\} \\ &= \sigma^{-2}\Big(\gamma(i-j) - \sum_{r=1}^{p} \phi_r \gamma(r - |i-j|)\Big). \end{aligned}$$

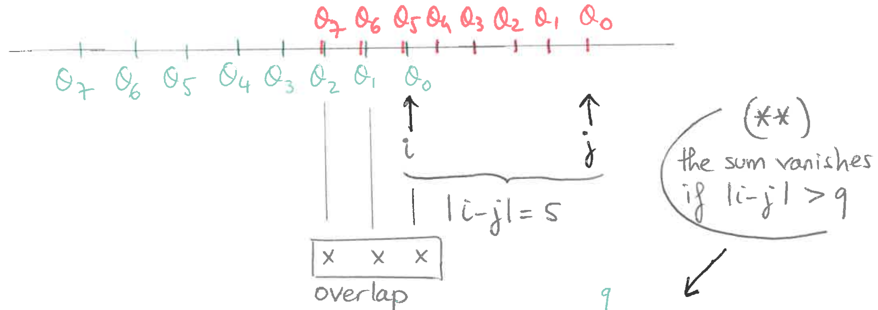


- $\min(i,j) > m$.

$$\begin{aligned} \mathbb{E}\, W_i W_j &= \sigma^{-2}\,\mathbb{E}\,(X_i - \phi_1 X_{i-1} - \cdots - \phi_p X_{i-p}) \\ &\qquad \times (X_j - \phi_1 X_{j-1} - \cdots - \phi_p X_{j-p}) \\ &= \sigma^{-2}\,\mathbb{E}\,(Z_i + \theta_1 Z_{i-1} + \cdots + \theta_q Z_{i-q}) \\ &\qquad \times (Z_j + \theta_1 Z_{j-1} + \cdots + \theta_q Z_{j-q}). \end{aligned}$$

Ex:

Take $j > i$.




$$\theta_2\theta_7 + \theta_1\theta_6 + \theta_0\theta_5 = \sum_{r=0}^{q} \theta_r \theta_{r+|i-j|}$$

- Otherwise take e.g. $i > 2m$ and $1 \leq j \leq m$.

  Then $\mathbb{E}\, W_i W_j = \sigma^{-1}\,\mathbb{E}\,(Z_i + \theta_1 Z_{i-1} + \cdots + \theta_q Z_{i-q}) X_j$

  $= 0$ since the process is causal.

Summary:

$$K(i,j) = \begin{cases} \sigma^{-2}\gamma(i-j) & \text{if } 1 \leq i, j \leq m \\ \sigma^{-2}\Big(\gamma(i-j) - \sum_{r=1}^{p} \phi_r \gamma(r - |i-j|)\Big) & \text{if } • \\ \sum_{r=0}^{q} \theta_r \theta_{r+|i-j|} & \text{if } \times \\ 0 & \text{if } \circ \end{cases}$$

39

Applying the innovations algorithm to $W_t$, we observe that $\theta_{n,j}$ for $n \geq m$ and $j > q$. Indeed, expanding the expressions for $\theta_{n,n-k}$ in the algorithm page 35,

$$\theta_{n,n} = v_0^{-1}\, K(n+1, 1)$$
$$\theta_{n,n-1} = v_1^{-1}\left(K(n+1,2) - \theta_{1,1}\theta_{n,n} v_0\right)$$
$$\theta_{n,n-2} = v_2^{-1}\left(K(n+1,3) - \theta_{22}\theta_{nn} v_0 - \theta_{21}\theta_{n,n-1} v_1\right)$$
$$\vdots$$
$$\theta_{n,q+1} = v_{n-q-1}^{-1}\left(\underbrace{K(n+1, n-q)}_{=0} - \underbrace{\text{previous terms}}_{=0}\right)$$

All these vanishe since $K(\cdot,\cdot) = 0$ here, see (**) p. 38

$$\theta_{n,q} = v_{n-q}^{-1}\left(\underbrace{K(n-1, n-q+1)}_{\text{non-zero}} - \underbrace{\text{previous terms}}_{=0}\right) \Rightarrow \theta_{n,q} > 0$$
$$\vdots$$

We see that

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^{n} \theta_{n,j}\left(W_{n+1-j} - \hat{W}_{n+1-j}\right), & 1 \leq n < m \\ \hat{W}_{n+1} = \sum_{j=1}^{q} \theta_{n,j}\left(W_{n+1-j} - \hat{W}_{n+1-j}\right), & n \geq m \\ r_n := E\left(W_{n+1} - \hat{W}_{n+1}\right)^2 \end{cases}$$

coefficients $\theta_{n,j}$ and $r_n$ found recursively by applying the innovation algorithm to $W_t$, with $K(i,j)$ as derived on pages 37, 38.

Next, notice that

$X_n$ $(n \geq 1)$ = linear combination of $W_1, \dots, W_n$

&

$W_n$ $(n \geq 1)$ = linear combination of $X_1, \dots, X_n$

$\Rightarrow$ Best linear predictor of $Y$ in terms of $\{X_1, \dots, X_n\}$ = Best linear predictor of $Y$ in terms of $\{W_1, \dots, W_n\}$.

40

Thus $\hat{W}_{n+1} = P_n W_{n+1}$ and $\hat{X}_{n+1} = P_n X_{n+1}$

same operator $P_n$

Now, projecting each side of
$$\begin{cases} W_t = \sigma^{-1} X_t, & t = 1, \dots, m \\ W_t = \sigma^{-1}\phi(B) X_t, & t > m \end{cases}$$
onto $\mathcal{H}_{t-1}$, we obtain

$$\begin{cases} \hat{W}_t = \sigma^{-1}\hat{X}_t, & t = 1, \dots, m \\ \hat{W}_t = \sigma^{-1}\left(\hat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}\right), & t > m \end{cases}$$

From which we get $X_t - \hat{X}_t = \sigma\left(W_t - \hat{W}_t\right)$. $\forall t$.

Plugging this back into the expressions of $\hat{W}_{n+1}$ page 39, we conclude that

$$\boxed{\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{n,j}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & 1 \leq n < m \\ \underbrace{\phi_1 X_n + \dots + \phi_p X_{n+1-p}}_{\text{AR part}} + \underbrace{\sum_{j=1}^{q} \theta_{n,j}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right)}_{\text{MA part}}, & n \geq m \end{cases}}$$

& $\mathbb{E}\left(X_{n+1} - \hat{X}_{n+1}\right)^2 = \sigma^2\, \mathbb{E}\left(W_{n+1} - \hat{W}_{n+1}\right)^2 = \sigma^2 r_n$,

where $r_n$ and $\theta_{n,j}$ are computed from the innovation algorithm page 35, and $K(i,j)$ is given on page 38.

Remarks: (i) Requires the storage of at most $p$ past observations $X_n, \dots, X_{n+1-p}$, and at most $q$ past innovations $\left(X_{n+1-j} - \hat{X}_{n+1-j}\right)$, $j = 1, \dots, q$.

(ii) AR(p) : $\hat{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p}$ (special case)

## • h-step prediction of ARMA(p,q), $h \geq 1$.

Recall the expression of the h-step ahead predictor appearing at the bottom of page 35, applied to the process $\{W_t\}$:

$$P_{\mathcal{H}_n} W_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left( W_{n+h-j} - \hat{W}_{n+h-j} \right)$$

$$= \sigma^{-1} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left( X_{n+h-j} - \hat{X}_{n+h-j} \right)$$

In addition, $W_t = \sigma^{-1} X_t$, $t = 1, \ldots, m$

$W_t = \sigma^{-1} \phi(B) X_t$, $t > m$

Applying the operator $P_{\mathcal{H}_n}$ on both sides yields

$$\begin{cases} P_{\mathcal{H}_n} W_{n+h} = \sigma^{-1} P_{\mathcal{H}_n} X_{n+h}, & \text{if } n+h \leq m \\ P_{\mathcal{H}_n} W_{n+h} = \sigma^{-1} \underbrace{P_{\mathcal{H}_n} \phi(B) X_{n+h}}_{= P_{\mathcal{H}_n}\left( X_{n+h} - \phi_1 X_{n+h-1} - \cdots - \phi_p X_{n+h-p} \right)}, & \text{if } n+h > m \end{cases}$$

We obtain

$$\boxed{\begin{aligned} P_{\mathcal{H}_n} X_{n+h} &= \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left( X_{n+h-j} - \hat{X}_{n+h-j} \right), \quad \text{if } 1 \leq h \leq m-n \\ P_{\mathcal{H}_n} X_{n+h} &= \sum_{i=1}^{p} \phi_i P_{\mathcal{H}_n} X_{n+h-i} + \sum_{j=h}^{q} \theta_{n+h-1,q} \left( X_{n+h-j} - \hat{X}_{n+h-j} \right) \quad \text{if } h > m-n \end{aligned}}$$

Once $\hat{X}_1, \ldots, \hat{X}_n$ are computed, n fixed, we can iteratively determine $P_{\mathcal{H}_n} X_{n+1}$, $P_{\mathcal{H}_n} X_{n+2}$, $P_{\mathcal{H}_n} X_{n+3}, \ldots$



## IV. THE PARTIAL AUTOCOVARIANCE (PACF) FUNCTION

The partial autocovariance function (PACF) of a stationary process is defined as

$$\phi_{11} = \mathrm{Corr}(X_1, X_0) = \rho(1)$$

$$\phi_{hh} = \mathrm{Corr}\left( X_{h+1} - P_{K_1} X_{h+1},\ X_1 - P_{K_1} X_1 \right)$$

$K_1 = \overline{sp}\{X_2, \ldots, X_n\}$.

= correlation between $X_{h+1}$ and $X_1$, with the linear effect of $X_2, \ldots, X_h$ removed (remove the effect of correlations due to the terms at shorter lags).

On page 31, we saw that the last coefficient $\phi_{n,n}$ of $\underline{\phi}_n$, satisfying $\Gamma_n \underline{\phi}_n = \underline{\gamma}_n$ (page 28), can be expressed

$$\phi_{n,n} = \frac{\langle X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\| X_1 - P_{K_1} X_1 \|}$$

$$= \frac{\langle X_{n+1} - P_{K_1} X_{n+1},\ X_1 - P_{K_1} X_1 \rangle}{\sqrt{\| X_{n+1} - P_{K_1} X_{n+1} \| \, \| X_1 - P_{K_1} X_1 \|}}$$

Since those two terms are equal

$$= \mathrm{Corr}\left( X_{n+1} - P_{K_1} X_{n+1},\ X_1 - P_{K_1} X_1 \right)$$

$\Rightarrow$ The PACF coefficient $\phi_{nn}$ corresponds to the last term in $\Gamma_n \underline{\phi}_n = \underline{\gamma}_n$, which can be efficiently computed using the Durbin-Levinson algorithm.

Ex: • AR(1) model $X_t = \phi X_{t-1} + Z_t$, $|\phi|<1$

• $\Gamma_1 \phi_1 = \gamma_1 \iff \gamma(0)\phi_{11} = \gamma(1) = \phi\gamma(0)$

(expression of ACF $\gamma$ of AR(1) process given on p. 6)

$$\boxed{\phi_{11} = \phi}$$

• $\Gamma_2 \phi_2 = \gamma_2 \iff \gamma(0)\begin{pmatrix}1&\phi\\ \phi&1\end{pmatrix}\begin{pmatrix}\phi_{2,1}\\ \phi_{2,2}\end{pmatrix} = \begin{pmatrix}\gamma(1)\\ \gamma(2)\end{pmatrix} = \gamma(0)\begin{pmatrix}\phi\\ \phi^2\end{pmatrix}$

yields $\phi_{2,1} = \phi$ and $\boxed{\phi_{2,2} = 0}$

And we see that $\phi_{n,n} = 0 \quad \forall n \geq 2$.

↳ the PACF of an AR(1) process vanishes at lags $\geq 2$

(Alternatively, note that $P_{K_1} X_3 = \phi X_2 = P_{K_1} X_1$,
($K_1 = \overline{sp}\{X_2\}$)

so that

$$\begin{aligned} corr(X_3 - P_{K_1}X_3,\ X_1 - P_{K_1}X_1) &= corr(\underbrace{X_3 - \phi X_2}_{= Z_3},\ X_1 - \phi X_2) \\ &= corr(Z_3,\ X_1 - \phi X_2) \\ &= 0 \quad \text{(causality)}. \end{aligned}$$

∎

• AR(p) $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t$, $Z_t \sim WN(0,\sigma^2)$.

We saw on page 40 that $\hat{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p}$,
$\forall n \geq p$.



In particular,

$$\hat{X}_{p+1} = \phi_1 X_p + \dots + \overset{=\phi_{pp}}{\boxed{\phi_p}} X_1$$

$$\hat{X}_{p+2} = \phi_1 X_{p+1} + \dots + \phi_p X_2 + \overset{=\phi_{p+1,p+1}}{\boxed{0}} X_1$$

$$\hat{X}_{p+3} = \phi_1 X_{p+2} + \dots + \phi_p X_3 + 0 X_2 + \overset{=\phi_{p+2,p+2}}{\boxed{0}} X_1$$

.../...

⇒ The PACF of an AR(p) process vanishes at lags $> p$

On the other hand, its ACF does not vanish at any lag. ∎

• MA(q) $X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$, invertible
$= \theta(B) Z_t$.

Let $\psi(z) = \sum_{i \geq 0} \pi_i z^i$ the power serie expansion of $\frac{1}{\theta(z)}$.

Then $Z_t = X_t + \sum_{i\geq 1} \pi_i X_{t-i}$ ($\pi_0 = 1$),

and $X_t = Z_t - \sum_{i \geq 1} \pi_i X_{t-i} = AR(\infty)$.

⇒ Since an MA(q) process can be represented as an AR(∞) time series, we see that its PACF will never cut-off.

Indeed, $\hat{X}_{n+1} = P_{\mathcal{H}_n} X_{n+1}$

$$= P_{\mathcal{H}_n}\left(Z_{n+1} - \sum_{i\geq 1} \pi_i X_{n+1-i}\right)$$

$$= -\sum_{i=1}^{n} \pi_i X_{n+1-i} - \sum_{i \geq n+1} \pi_i P_{\mathcal{H}_n} X_{n+1-i}.$$

∎

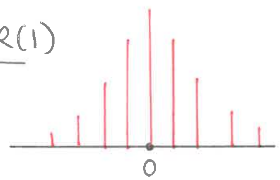
Summary:



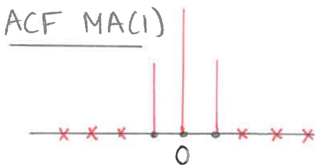
| Model | ACF | PACF |
|---|---|---|
| AR(p) | decays | zero for $h > p$ |
| MA(q) | zero for $h > q$ | decays |
| ARMA(p, q) | decays | decays |

To see this, proceed as for MA(q).
⇒ Can be used to get an idea of the order of magnitude of p and q when fitting an ARMA process to real data.

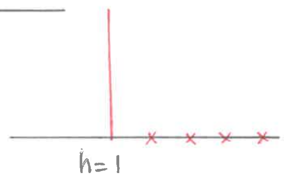
ACF AR(1)



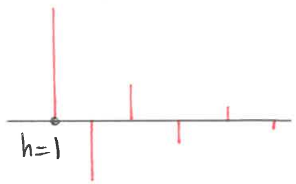

ACF MA(1)




PACF AR(1)




PACF MA(1)




We can show that for an MA(1) process $X_t = Z_t + \theta Z_{t-1}$,

$$\phi_{h,h} = -\frac{(-\theta)^h}{1+\theta+\cdots+\theta^{2h}}$$

# V. ESTIMATION FOR ARMA MODELS



## V.1. Maximum Likelihood Estimation.

We want to estimate the parameters $\phi_1, .., \phi_p, \theta_1, .., \theta_q, \sigma^2$ of an ARMA(p, q) model, assuming (for now) that

→ The order p, q is known

→ The process is zero-mean

If not, simply substract the sample mean $\bar{y}$ of $\{Y_t\}$, fit an ARMA model $\phi(B) X_t = \theta(B) Z_t$ to $X_t = Y_t - \bar{y}$, and then use $X_t + \bar{y}$ as a model for $Y_t$.

Assume that $Z_t = \text{iid } \mathcal{N}(0, \sigma^2)$. The process $\{X_t\}$ is then Gaussian, and a usual approach is to select $\phi = (\phi_1, .., \phi_p)$, $\theta = (\theta_1, .., \theta_q)$ and $\sigma^2$ maximizing the likelihood $L(\phi, \theta, \sigma^2) = f(x_1, .., x_n \mid \phi, \theta, \sigma^2)$, where $f(\cdot \mid \phi, \theta, \sigma^2)$ denotes the joint Gaussian density of the given ARMA(p, q) model. It is given by

$$L(\phi, \theta, \sigma^2) = \frac{1}{(2\pi)^{n/2} |\underline{\Gamma}_n|} \exp\left\{-\frac{1}{2} \underline{X}_n^t \underline{\Gamma}_n \underline{X}_n\right\}, \quad (*)$$

where $\underline{\Gamma}_n$ denotes the covariance matrix of $\underline{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$.

(its expression is given page 28)

The maximum likelihood estimates are asymptotically efficient ( lowest variance amongst the class of asymp. unbiased estimators )(see MS: MLE ). In addition, it is possible to show that even if the Gaussian assumption fails to hold, the asymptotic distribution of the MLE is the same as in the Gaussian case. (47)

$\Rightarrow$ It makes sense to consider the Gaussian likelihood of an ARMA(p,q) process for parameter estimation.

However, it usually is a difficult optimization problem due to the presence of many local minima. Usually performed numerically & recursively $\to$ we need to choose a good starting point. We may use (i) Yule-Walker for an AR(p) process (ii) the Innovations algorithm for an MA(q) process and (iii) Hannan-Rissanen for a general ARMA(p,q) process. We return to these algorithms in Sections V.2/3/4, respectively.

Note that we may write $\underline{X}_n = \underline{\underline{C}}_n ( \underline{X}_n - \hat{\underline{X}}_n )$, where the $(n\times n)$ matrix $\underline{\underline{C}}_n$ is lower triangular, whose coefficients are obtain from the innovations algorithm,

$$\underset{(n\times n)}{\underline{\underline{C}}_n} = \begin{pmatrix} 1 & 0 & 0 & & 0 \\ \theta_{11} & 1 & 0 & & 0 \\ \theta_{22} & \theta_{21} & 1 & \ddots & 0 \\ \vdots & & \vdots & & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{pmatrix},$$

since $\hat{X}_j = \sum_{\ell=1}^{j-1} \theta_{j-1,\ell} ( X_{j-\ell} - \hat{X}_{j-\ell} )$, $j = 2, \dots, n$,

and $\hat{X}_1 = 0$.

The advantage of expressing $\underline{X}_n$ in terms of $(\underline{X}_n - \hat{\underline{X}}_n)$ is that $(\underline{X}_n - \hat{\underline{X}}_n)$ has a diagonal covariance matrix (48)

$$\underset{(n\times n)}{\underline{\underline{D}}_n} = \begin{pmatrix} v_0 & & & 0 \\ & v_1 & & \\ & & \ddots & \\ 0 & & & v_{n-1} \end{pmatrix},$$ whose terms are also obtained from the innovations algorithm.

$\Rightarrow$ $\underline{\underline{\Gamma}}_n = \underline{\underline{C}}_n \underline{\underline{D}}_n \underline{\underline{C}}_n^t$, and

$$\underline{X}_n^t \underline{\underline{\Gamma}}_n^{-1} \underline{X}_n = ( \underline{X}_n - \hat{\underline{X}}_n )^t \underline{\underline{D}}_n^{-1} ( \underline{X}_n - \hat{\underline{X}}_n )^t = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{v_{j-1}}.$$

In addition, $|\underline{\underline{\Gamma}}_n| = |\underline{\underline{C}}_n|^2 |\underline{\underline{D}}_n| = v_0 \times \dots \times v_{n-1}$.

The likelihood $(*)$ page 46 reduces to

$$\boxed{L(\phi, \theta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^r r_0 \times \dots \times r_{n-1}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \right\}},$$

where $r_\ell := v_\ell / \sigma^2$.

The log-likelihood is

$$\ell(\phi, \theta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=0}^{n-1} \log r_i - \frac{1}{2\sigma^2} \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}.$$

Differentiating with respect to $\sigma^2$ shows that the MLE $\hat{\sigma}^2$ of $\sigma^2$ satisfies

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{\phi}, \hat{\theta}), \text{ with } S(\hat{\phi}, \hat{\theta}) := \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}},$$

and $\hat{\phi}, \hat{\theta}$ minimize $\left\{ \log n^{-1} S(\phi, \theta) + \frac{1}{n} \sum_{j=1}^{n} \log r_j \right\}$.

The minimization is done numerically, and can be done under further simplifications: drop the $\log r_{j-1}$ terms (since $v_j \to \sigma^2$, $r_j \to 1$ and $\frac{1}{n}\sum \log r_j \to 0$), or approximate the $\hat{X}_j$ terms for example.

For an ARMA(p,q) process, the MLE is asymptotically normally distributed,

$$\begin{pmatrix}\hat\phi\\ \hat\theta\end{pmatrix} - \begin{pmatrix}\phi\\ \theta\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{n}\begin{pmatrix}\Gamma_{\phi\phi} & \Gamma_{\phi\theta}\\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta}\end{pmatrix}^{-1}\right),$$

where

$$\begin{pmatrix}\Gamma_{\phi\phi} & \Gamma_{\phi\theta}\\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta}\end{pmatrix} = \Sigma_{(X,Y)}, \quad \begin{aligned} X &= (X_1, \dots, X_p)^t \\ Y &= (Y_1, \dots, Y_q)^t \\ \phi(B) X_t &= Z_t \\ \theta(B) Y_t &= Z_t \end{aligned}$$

↳ can construct confidence intervals.

× Remark: In some simple cases, the maximization of the log-likelihood can be done explicitely.

→ AR(1) process $X_t = \phi X_{t-1} + Z_t \quad Z_t \sim WN(0, \sigma^2)$

A usual linear regression problem, where $X_{t-1}$ is the predictor and $X_t$ the response variable. Under the Gaussian assumption, the MLE coincides with the LS estimate $\hat\phi = (\underline{X}^t \underline{X})^{-1} \underline{X}^t \underline{y}$, where

$$\underset{(n-1)\times 1}{\underline{X}} = \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1}\end{pmatrix} = \text{matrix of observations}, \quad \underset{(n-1)\times 1}{\underline{y}} = \begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix},$$



so that

$$\hat\phi = \frac{\sum_{t=1}^{n-1} x_t x_{t+1}}{\sum_{t=1}^{n-1} x_t^2}, \quad \text{and } \hat\sigma^2 = \frac{1}{n-1}\sum_{t=2}^{n} (x_t - \hat\phi x_{t-1})^2.$$

Likewise, for an AR(p) process, we obtain

$$\hat{\underline{\phi}} = \begin{pmatrix}\hat\phi_1 \\ \vdots \\ \hat\phi_p\end{pmatrix} = \left(\sum_{t=p+1}^{n} \underline{x}_t \underline{x}_t^t\right)^{-1} \left(\sum_{t=p+1}^{n} \underline{x}_t \underline{x}_t^t\right), \quad \underline{x}_t = \begin{pmatrix} x_{t-1} \\ \vdots \\ x_{t-p}\end{pmatrix}.$$

→ MA(1) process $X_t = Z_t + \theta Z_{t-1}$.

- $\left.\begin{aligned} X_2 &= Z_2 + \theta Z_1 \\ Z_1 &= X_1 - \theta Z_0 \end{aligned}\right\} \to X_2 = \theta X_1 + Z_2 - \theta^2 Z_0$

  $\Rightarrow E\, X_2 | X_1 = \theta X_1$ (assuming $E(Z_0 | X_1) = 0$)

  & $\operatorname{Var} X_2 | X_1 = \sigma^2 (1 + \theta^4)$

Therefore, when decomposing the joint likelihood as a product of marginals, the first term $f(x_2 | x_1)$ already includes non linearities in the parameters $\theta$ and $\sigma^2$. The same applies to all the terms $f(x_t | x_1, \dots, x_{t-1})$.

⇒ Direct maximization is not tractable. The moving average part is causing the trouble in likelihood maximization.

## IV2. Yule-Walker estimation.

The Yule-Walker (YW) procedure for parameter estimation is particularly adapted to pure auto-regressive models $X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t$ (and used as initialization for likelihood maximization).

Multiplying the expression of an AR(p) on both sides by $X_{t-j}$, $j=0,\dots,p$, and taking expectations,

$$E(X_t X_{t-j}) - \phi_1 E(X_{t-1} X_{t-j}) - \phi_2 E(X_{t-2} X_{t-j}) - \dots - \phi_p E(X_{t-p} X_{t-j}) = E(X_{t-j} Z_t)$$

$$\begin{array}{ll} j=0 & \gamma(0) - \phi_1 \gamma(1) - \dots - \phi_p \gamma(p) = \sigma^2 \\ j=1 & \gamma(1) - \phi_1 \gamma(0) - \dots - \phi_p \gamma(p-1) = 0 \\ & \vdots \\ j=p & \gamma(p) - \phi_1 \gamma(p-1) - \dots - \phi_p \gamma(0) = 0 \end{array}$$

$$\Leftrightarrow \underbrace{\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & \vdots & & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{pmatrix}}_{=\underline{\Gamma}_p} \underbrace{\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}}_{=\underline{\phi}_p} = \underbrace{\begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix}}_{=\underline{\gamma}_p}$$

$$\& \quad \sigma^2 = \gamma(0) - \underline{\phi}_p^t \underline{\gamma}_p$$

$$\boxed{\underline{\Gamma}_p \underline{\phi}_p = \underline{\gamma}_p}$$

We have already encounted this system of linear equations in the context of 1-step ahead prediction



To get an estimate of $\underline{\phi}_p$, replace the covariance terms by their sample estimates $\hat{\gamma}$.

$$\boxed{\text{Yule-Walker equations for } \hat{\underline{\phi}}_p : \begin{cases} \hat{\underline{\Gamma}}_p \hat{\underline{\phi}}_p = \hat{\underline{\gamma}}_p \\ \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\underline{\phi}}^t \hat{\underline{\gamma}}_p \end{cases}}$$

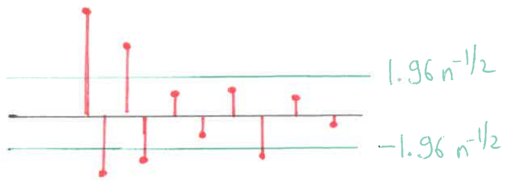
Can use the Durbin-Levinson algorithm.

If $\hat{\gamma}(0) > 0$, then $\hat{\underline{\Gamma}}_p$ is necessarily non-singular, and it is possible to show that the fitted AR(p) model is causal.

Asymptotic properties of $\hat{\underline{\phi}}_p$ can be established as well:

$$n^{1/2}(\hat{\underline{\phi}}_p - \underline{\phi}_p) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \underline{\Gamma}_p^{-1}), \quad \hat{\sigma}^2 \xrightarrow{P} \sigma^2$$

x Remark: If $\{X_t\}$ is AR(p), and we fit an auto-regressive model of order $m > p$ using the YW equations $\hat{\underline{\phi}}_m = \hat{\underline{\Gamma}}_m^{-1} \hat{\underline{\gamma}}_m$, then it is possible to show that the last component of $\hat{\underline{\phi}}_m$ (aka $\hat{\phi}_{mm}$, the estimate of the PACF at lag m, equal to zero) is asymptotically $\mathcal{N}(0, n^{-1})$. Therefore, if AR(p) is appropriate, then all the $\hat{\phi}_{mm}$, $m > p$ should be compatible with an $\mathcal{N}(0, n^{-1})$ distribution, i.e. $\hat{\phi}_{mm} \in [-1.96 n^{-1/2}, 1.96 n^{-1/2}]$.



## V.3. Innovation's algorithm for MA(q).



Recall the 1-step ahead prediction of $X_{n+1}$ using $X_1,\dots,X_n$:

$$\hat X_{n+1} = \theta_{n,1}\boxed{(X_n-\hat X_n)} + \theta_{n,2}\boxed{(X_{n-1}-\hat X_{n-1})} + \dots + \theta_{n,n}\boxed{(X_1-\hat X_1)}$$

residuals $\equiv$ noise $Z_j$

& $E(X_{n+1}-\hat X_{n+1})^2 = v_n \equiv$ residual noise variance.

$\Rightarrow$ The idea is to fit an innovation's MA(m) model

$$X_t = Z_t + \hat\theta_{m,1} Z_{t-1} + \dots + \hat\theta_{m,m} Z_{t-m}, \quad Z_t \sim WN(0,\hat v_t)$$

obtained from the innovation's algorithm with the ACF replaced with the sample ACF.

× Remark: For an MA(q) process, the estimator $\hat\theta_q = (\theta_{q,1},\dots,\theta_{qq})$ is not necessarily a consistent estimator of $\theta_q = (\theta_1,\dots,\theta_q)$. To get consistency, it is necessary to use the estimators $(\theta_{m,1},\dots,\theta_{m,q})$, with $m = m(n)$ satisfying $m(n)\to+\infty$, and $n^{-1/3}m(n)\to 0$. A central limit theorem holds under the same conditions.

• For an MA(q) process, $\gamma(m)=0$ for $m>q$. Moreover, we know that the sample ACF $\hat\rho(m)$ $(m>q)$ is approximately $\mathcal{N}(\rho(m)=0,\ n^{-1}\{1+2\rho^2(1)+\dots+2\rho^2(q)\})$. This can be used to pre-select the order of the MA(q) process (construct confidence bands).

## V.4. Hannan-Rissanen Algorithm.



A pure auto-regressive model is fitted like a linear regression model with coefficients $\phi_1,\dots,\phi_p$. When $q>0$, the noise terms are unobserved, but can be estimated and then used to compute the least squares estimates of $\phi_1,\dots,\phi_p,\theta_1,\dots,\theta_q$. Indeed, for an invertible ARMA(p,q) process, the representation $Z_t = \sum_{j\geq 0}\alpha_j X_{t-j}$ suggests to estimate $Z_t$ as the residual of a fitted high order AR(m) process:

$$\hat Z_t = X_t - \hat\phi_{m1}X_{t-1} - \dots - \hat\phi_{mm}X_{t-m}, \quad t=m+1,\dots,n \quad (m>\max(p,q)).$$

Then, compute the LS estimate of $X_t$ onto $(X_{t-1},\dots,X_{t-p},\hat Z_{t-1},\dots,\hat Z_{t-q})$:

$$\underbrace{\begin{pmatrix} X_{m+1+q} \\ \vdots \\ X_n \end{pmatrix}}_{\text{target } \underline{Y}} \text{ on } \underbrace{\begin{pmatrix} X_{m+q} & \dots & X_{m+1+q-p} & \hat Z_{m+q} & \dots & \hat Z_{m+1} \\ & & & & & \\ X_{n-1} & \dots & X_{n-p} & \hat Z_{n-1} & \dots & \hat Z_{n-q} \end{pmatrix}}_{\text{matrix of observations } \underline{\underline{Z}}}$$

Then $\hat\beta := (\hat\phi_1,\dots,\hat\phi_p,\hat\theta_1,\dots,\hat\theta_q)^t = (\underline{\underline{Z}}^t\underline{\underline{Z}})^{-1}\underline{\underline{Z}}^t\underline{Y}$.

## V.5. AIC / BIC

To select a model amongst a class of candidates with different values of $(p,q)$, we may consider a penalized criterion such as AIC or BIC (see SL: MODEL SELECTION).

AIC, penalty is $2(p+q+1)n/(n-p-q-2)$.

# VI - SPECTRAL ANALYSIS



A square integrable function $x(t)$ defined on $[0,T]$ admits a decomposition in terms of sinusoids

$$x(t) = \sum_{n=0}^{+\infty} \left\{ a_n \cos\left(2\pi \frac{n}{T} t\right) + b_n \sin\left(2\pi \frac{n}{T} t\right) \right\},$$

$$a_0 = \frac{1}{T}\int_0^T x(t)\, dt, \quad a_n = \frac{2}{T}\int_0^T x(t) \cos \frac{2\pi n t}{T}\, dt, \quad b_n = \frac{2}{T}\int_0^T x(t) \sin(..)\, dt$$

or, equivalently, in terms of the complex exponential

$$\boxed{x(t) = \sum_{k=-\infty}^{+\infty} c_k\, e^{2i\pi \frac{k}{T} t}}, \qquad c_k = \frac{1}{T}\int_0^T x(t)\, e^{-2i\pi \frac{k}{T} t}\, dt$$

contribution at frequency $\nu_k := \frac{k}{T}$ (discrete).

Likewise, for a function defined on the whole real line, and such that it belongs to $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, we obtain a similar decomposition

$$\boxed{x(t) = \int_{\mathbb{R}} \hat{x}(\nu)\, e^{2i\pi \nu t}\, d\nu}, \qquad \hat{x}(\nu) = \int_{\mathbb{R}} x(t)\, e^{-2i\pi \nu t}\, dt$$

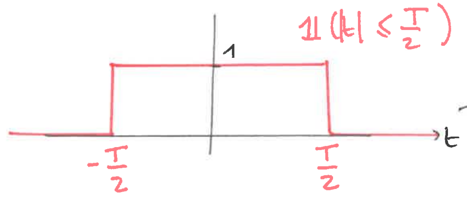
contribution of the complex exponential at frequency $\nu \in \mathbb{R}$.

$=$ Fourier Transform (FT) of $x(t)$ $=: \mathcal{F}(x(t))$.

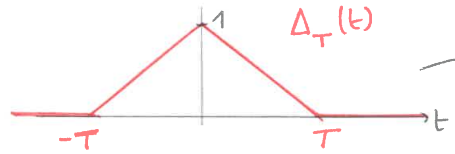
$\Rightarrow$ The quantity $S(\nu) := |\hat{x}(\nu)|^2$ plays a central role when analysing the frequency composition of a signal $x(t)$.



**Ex:**




$\longrightarrow \text{FT} = T\,\dfrac{\sin \pi T \nu}{\pi T \nu}$.




$\longrightarrow \text{FT} = T\left(\operatorname{sinc} \pi T \nu\right)^2$

For a (random) stationary process $\{X(t)\}$, we consider more generally the Power Spectral Density (PSD)

$$S_X(\nu) := \lim_{T\to\infty} \frac{1}{T} |\hat{X}_T(\nu)|^2, \quad \leftarrow$$

Signal continues over an $\infty$ time with same statistical properties: truncate

where $\hat{X}_T(\nu) = \mathcal{F}(X_T(t))$; $X_T(t) = X(t)\, \mathbb{1}\left(|t| \leq \frac{T}{2}\right)$.

The WIENER-KHINCHIN theorem states that for a stationary time-series, the PSD of $\{X(t)\}$ is equal to the Fourier Transform of its ACF. More formally, if $t$ is continuous, then

$$S_X(\nu) = \hat{R}_X(\nu) = \int_{\mathbb{R}} R(\tau)\, e^{-2i\pi\nu\tau}\, d\tau,$$

where $R(\tau) := \mathbb{E}\left(X(t)\, X(t-\tau)\right)$

$$= \int_{\mathbb{R}} S_X(\nu)\, e^{2i\pi\nu\tau}\, d\nu$$

Under integrability assumptions of $R(\tau)$

If $t$ is discrete (ex: ARMA(p,q) process), then

$$S_X(\nu) = \sum_{k\in\mathbb{Z}} \gamma(k)\, e^{-2i\pi\nu k},$$

where $\gamma(k) = ACF = \int_{-\frac{1}{2}}^{\frac{1}{2}} \underbrace{S_X(\nu)\, e^{2i\pi k\nu}\, d\nu}_{= dF(\nu)}, \quad k\in\mathbb{Z}.$

An ACF nec. has the representation $\int e^{2i\pi k\nu} dF(\nu)$ (Herglotz), $F$ = right-cnts, non-decrea, bounded

Under integrability condition $\sum_{k\in\mathbb{Z}} |\gamma(k)| < +\infty$.

Properties of $S_X(\nu)$.

(i) $S_X$ is periodic, with period 1.

(which holds since $e^{-2i\pi k\nu}$ is 1-periodic).

For this reason, we can restrict $\nu$ to the interval $[-\frac{1}{2}, \frac{1}{2}]$.

(ii) $S_X$ is even

(iii) $S_X(\nu) \geq 0$

## VI.1. PSD of an ARMA(p,q)

- WN$(0,\sigma^2)$ $\quad \gamma(k) = 0 \quad \forall k \neq 0, \quad S(\nu) = \gamma(0) = \sigma^2.$
- AR(1) $\quad X_t - \phi X_{t-1} = Z_t, \quad Z_t \sim WN(0,\sigma^2)$

We have seen that $\gamma(k) = \frac{\sigma^2}{1-\phi^2}\phi^{|k|}$ (page 6)

Thus

$$S_X(\nu) = \frac{\sigma^2}{1-\phi^2} \sum_k \phi^{|k|} e^{-2i\pi\nu k}$$

$$= \frac{\sigma^2}{1-\phi^2}\left\{1 + \sum_{k\geq 1} \phi^k \left(e^{-2i\pi\nu k} + e^{2i\pi\nu k}\right)\right\}$$



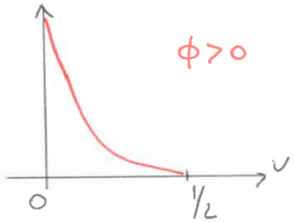
$$S_X(\nu) = \frac{\sigma^2}{1-\phi^2}\left\{1 + \frac{\phi e^{-2i\pi\nu}}{1-\phi e^{-2i\pi\nu}} + \frac{\phi e^{2i\pi\nu}}{1-\phi e^{2i\pi\nu}}\right\}$$

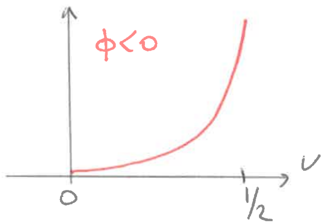
$$= \frac{\sigma^2}{1-\phi^2} \frac{1-\phi e^{-2i\pi\nu} e^{2i\pi\nu}}{(1-\phi e^{-2i\pi\nu})(1-\phi e^{2i\pi\nu})}$$

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

$\Rightarrow$ If $\phi > 0$, the spectrum is dominated by low frequency components (smooth in time domain)

If $\phi < 0$, ——— " ——— by high frequency ——— " ——— (rough in time domain)

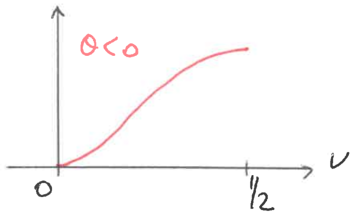



- MA(1) $\quad X_t = Z_t + \theta Z_{t-1}, \quad Z_t \sim WN(0,\sigma^2).$

Then $\gamma(k) = \begin{cases} \sigma^2(1+\theta^2) & \text{if } k=0 \\ \sigma^2\theta & \text{if } |k|=1 \\ 0 & \text{o/w} \end{cases}$, (page 5)

$$S_X(\nu) = \sum_{k=-1}^{1} \gamma(k) e^{-2i\pi k\nu} = \sigma^2\left(1+\theta^2 + 2\theta\cos(2\pi\nu)\right).$$

59

- More generally, for a linear process $\{X_t\}$ such that $X_t = \sum_{k\geq 0} \psi_k Z_{t-k} = \psi(B) Z_t$, the PSD of $\{X_t\}$ is given by $S_X(\nu) = \sigma^2 |\psi(e^{-2i\pi\nu})|^2$.

For an ARMA$(p,q)$ process, $\psi(B) = \dfrac{\theta(B)}{\phi(B)}$, so

$$\boxed{S_X(\nu) = \sigma^2 \left| \frac{\theta(e^{-2i\pi\nu})}{\phi(e^{-2i\pi\nu})} \right|^2} = \text{rational spectrum}.$$

Using the factorization $\theta(z) = \theta_q (z - z_1) \dots (z - z_q)$
$\phi(z) = \phi_p (z - p_1) \dots (z - p_p)$,

$z_i$ = zeros
$p_j$ = poles,

we have
$$\boxed{S_X(\nu) = \sigma^2 \frac{\theta_q^2 \prod_{j=1}^{q} |e^{-2i\pi\nu} - z_j|^2}{\phi_p^2 \prod_{j=1}^{p} |e^{-2i\pi\nu} - p_j|^2}}$$

- For an AR(1), $\phi(z) = 1 - \phi z$, the pole is at $1/\phi$. If $\phi > 0$, the pole is to the right of $1$, so the spectral density decreases as $\nu$ moves away from 0. If $\phi < 0$, the pole is to the left of $-1$, and the spectral density achieves its maximum at $\nu = 1/2$.

- For an MA(1), $\theta(z) = 1 + \theta z$, the zero is at $-1/\theta$. If $\theta > 0$, the zero is to the left of $-1$, so the spectral density decreases as $\nu$ moves towards $-1$. If $\theta < 0$, the zero is to the right of $1$, the spectral density is at its minimum when $\nu = 0$.

60

- Remark: To get to the expression on page 59, note that for a linear process $X_t = \sum_{k\geq 0} \psi_k Z_{t-k} = \psi(B) Z_t$, the covariance function is $\gamma(k) = \sigma^2 \sum_{i\geq 0} \psi_i \psi_{i+h}$ (page 12).

Making use of the operator $\Gamma(B) := \sum_{k\in\mathbb{Z}} \gamma(k) B^k$, the PSD of $X_t$ can be written $S(\nu) = \sum_{k\in\mathbb{Z}} \gamma(k) e^{-2i\pi\nu k}$
$= \Gamma(e^{-2i\pi\nu})$

Note that
$$\Gamma(B) = \sigma^2 \sum_{k\in\mathbb{Z}} \sum_{i\geq 0} \psi_i \psi_{i+k} B^k$$
$$= \sigma^2 \sum_{i\geq 0} \psi_i \sum_{k\in\mathbb{Z}} \psi_{i+k} B^k$$
put $j = i+k$. then $\psi_j = 0$ for $j < 0$
$$= \sigma^2 \sum_{i\geq 0} \psi_i \sum_{j\geq 0} \psi_j B^{j-i}$$
$$= \sigma^2 \left(\sum_{i\geq 0} \psi_i B^{-i}\right)\left(\sum_{j\geq 0} \psi_j B^{j}\right) = \sigma^2 \psi(B^{-1}) \psi(B)$$

We get $S(\nu) = \Gamma(e^{-2i\pi\nu}) = \sigma^2 \psi(e^{-2i\pi\nu}) \psi(e^{2i\pi\nu})$
$$\boxed{S(\nu) = \sigma^2 |\psi(e^{2i\pi\nu})|^2}$$

## VI.2. Time-invariant Linear filters.

Filter = maps a time series $\{X_t\}$ to a time series $\{Y_t\}$

$$Y_t = \sum_{j\in\mathbb{Z}} \psi_j X_j$$ ← if $\psi_j = 0$ for $j < 0$, the filter is causal

time-invariant → independent of $t$.

Ex: • $Y_t = X_{-t}$ is linear but not time-invariant

• $Y_t = \frac{1}{2}(X_{t-1} + X_{t+1})$ is linear, time-invariant, not causal

• For polynomials $\phi(B)$, $\theta(B)$ with roots outside the unit circle, $\psi(B) = \theta(B)/\phi(B)$ is linear, time-invariant, causal

Result: Suppose that $\{X_t\}$ has spectral density $S_X(\nu)$, that $\psi$ is STABLE ($\sum|\psi_j| < \infty$). Then $Y_t = \psi(B)X_t$ has spectral density

$$S_Y(\nu) = \underbrace{|\psi(e^{2i\pi\nu})|^2}_{\text{power transfer function of the filter}} S_X(\nu).$$

The function $\nu \mapsto \psi(e^{2i\pi\nu})$ is called the frequency response or transfer function of the linear filter.

proof: The ACF of $\{Y_t\}$ is

$$\gamma_Y(h) = \mathbb{E}\left(\sum_{j\in\mathbb{Z}} \psi_j X_{t-j} \sum_{k\in\mathbb{Z}} \psi_k X_{t+h-k}\right)$$

$$= \sum_{j\in\mathbb{Z}} \psi_j \sum_{k\in\mathbb{Z}} \psi_k \, \mathbb{E} X_{t-j} X_{t+h-k}$$

$$= \sum_{j\in\mathbb{Z}} \psi_j \sum_{k\in\mathbb{Z}} \psi_k \, \gamma_X(h-k+j) \qquad \ell = h-k+j \quad (k = h-\ell+j)$$

$$= \sum_{j\in\mathbb{Z}} \psi_j \sum_{\ell\in\mathbb{Z}} \psi_{h-\ell+j} \, \gamma_X(\ell)$$

$\sum_j |\psi_j| < \infty$ and $\sum_\ell \gamma_X(\ell) < \infty$ $\Rightarrow$ $\sum_h |\gamma_Y(h)| < \infty$, so the spectral density of $\{Y_t\}$ is well defined.



$$S_Y(\nu) = \sum_{h\in\mathbb{Z}} \gamma(h) e^{-2i\pi\nu h}$$

$$= \sum_h \sum_j \psi_j \sum_\ell \psi_{h-\ell+j} \gamma_X(\ell) \, e^{-2i\pi\nu h}$$

$$= \sum_j \psi_j e^{2i\pi\nu j} \sum_\ell \gamma_X(\ell) e^{-2i\pi\nu\ell} \sum_h \psi_{h-\ell+j} e^{-2i\pi\nu(h+j-\ell)}$$

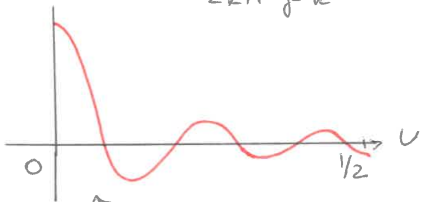
$$= \psi(e^{2i\pi\nu}) S_X(\nu) \psi(e^{-2i\pi\nu})$$

$$= |\psi(e^{2i\pi\nu})|^2 S_X(\nu).$$

Ex: (i) Moving average $Y_t = \frac{1}{2k+1} \sum_{j=-k}^{k} X_{t-j}$.

The transfer function is

$$\psi(e^{-2i\pi\nu}) = \frac{1}{2k+1} \sum_{j=-k}^{k} e^{-2i\pi j\nu} = \begin{cases} 1 & \nu = 0 \\ \dfrac{\sin(2\pi(k+\frac{1}{2})\nu)}{(2k+1)\sin\pi\nu} & \text{o/w} \end{cases}$$



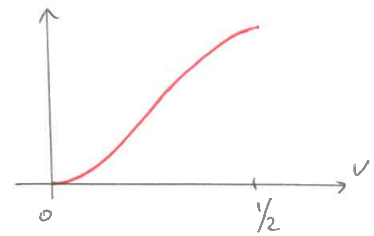

A LOW-PASS filter: preserves low frequencies, and diminishes high frequencies.

(ii) Differencing $Y_t = (1-B)X_t$ = time invariant, causal, linear filter

$$\psi(e^{-2i\pi\nu}) = 1 - e^{-2i\pi\nu}$$

$$|\psi(e^{-2i\pi\nu})|^2 = 2(1-\cos 2\pi\nu)$$

A HIGH PASS filter: preserves high frequences. Used to remove trends.



# VII - ARIMA PROCESSES.



ARMA processes are an important class of models for stationary time series. A generalisation of this class, allowing to model non-stationary time series, is provided by ARIMA and SARIMA processes.

Let $d \geq 0$. Then $\{X_t\}$ is an ARIMA$(p,d,q)$ process if $Y_t := (1-B)^d X_t$ is a causal ARMA$(p,q)$ process:

$$\phi(B) Y_t = \theta(B) X_t$$

$$\underbrace{\phi(B)(1-B)^d}_{} X_t = \theta(B) X_t$$

ARIMA$(p,d,q)$

Put $\phi^*(B) = \phi(B)(1-B)^d$.
Then $\phi^*$ has a zero of order $d$ at $z=1$ $\Rightarrow$ The process $\{X_t\}$ is stationary if and only if $d=0$.

$(1-B)^d$ = order-$d$ differentiation.
We mentioned in section I and V already that differentiating is used for trend removal. Indeed, the operator $(1-B)$ acts as a high pass filter, and attenuate low frequency ($\equiv$ trend) components.

→ Selecting $p$, $d$ and $q$. (Box-Jenkins method)

(i) Plot the time series, look for trends. A slowly decaying ACF is symptomatic of the presence of a trend.



(ii) Differentiate until the time series "look" stationary (→ select $d$)

(iii) Identify $(p,q)$ by looking at the sample ACF and PACF of $(1-B)^d X_t$ (→ select $p,q$)

(iv) Fit an ARMA$(p,q)$ model to $(1-B)^d X_t$ using maximum likelihood (section V)

(v) Check Residuals ("diagnostic checking")

(vi) Repeat for a range of suitable candidate values of $(p,d,q)$, and report the AIC / BIC.

↓

Use the fitted ARIMA$(p,d,q)$ for prediction.

→ h-step ahead prediction of ARIMA models.

- Notation: $Y_t = (1-B)^d X_t$ for $t = 1, 2, \ldots$

$d=1$ → use $X_0, X_1, X_2, \ldots$ to construct $Y_t$, $t \geq 1$

$Y_1 = X_1 - X_0$, $Y_2 = X_2 - X_1, \ldots$

$d=2$ → need $X_{-1}, X_0, X_1, X_2, \ldots$

$Y_1 = X_1 - 2X_0 + X_{-1}$, $Y_2 = \ldots / \ldots$

For a general $d$, we need $X_{1-d}, \ldots, X_0$, $(d \geq 0)$ to compute $Y_t$, $t \geq 1$.

- Assumption: $X_{1-d}, \ldots, X_0 \perp Y_t$, $\forall t \geq 1$

Since $(1-B)^d = \sum_{j=0}^{d} \binom{d}{j} (-1)^j B^j$, we can write

$$X_t = Y_t - \sum_{j=1}^{d} \binom{d}{j} (-1)^j X_{t-j}, \quad t \geq 1 \quad \text{—— } (*)$$

In the previous notation, we assume that we observe $\begin{pmatrix} X_{1-d}, X_{2-d}, \dots, X_0, X_1, \dots, X_n \\ Y_1, \dots, Y_n \end{pmatrix}$

- $\underline{\text{Goal}}$: best prediction of $X_{n+h}$ $(h \geq 1)$ based on $X_{1-d}, \dots, X_n$.

  Put $S_n := \overline{sp}\{X_{1-d}, \dots, X_n\}$.

  We are looking for $P_{S_n} X_{n+h}$.

  We write $P_n Y_{n+h} := P_{\overline{sp}\{Y_1, \dots, Y_n\}} Y_{n+h}$, and $\hat{Y}_{n+1} = P_n Y_{n+1}$.

  Under the assumption that $X_{1-d}, \dots, X_0$ are uncorrelated with $Y_t$, $t \geq 1$,

$$S_n = \overline{sp}\{X_{1-d}, \dots, X_0, X_1, \dots, X_n\}$$

since the $X_j$ / $Y_j$ are linear combinations of one another

$$= \overline{sp}\{\underbrace{X_{1-d}, \dots, X_0}, \underbrace{Y_1, \dots, Y_n}\} \quad (\perp)$$

$$\Rightarrow P_{S_n} Y_{n+h} = \underbrace{P_{S_0} Y_{n+h}}_{=0} + P_n Y_{n+h}$$

Apply $P_{S_n}$ on both side of $(*)$ p. 64.

$$\boxed{P_{S_n} X_{n+h} = P_n Y_{n+h} - \sum_{j=1}^{d} \binom{d}{j} (-1)^j P_{S_n} X_{n+h-j}.}$$

Can be easily computed since $\{Y_t\}$ is an ARMA$(p,q)$ process

→ Calculate $P_{S_n} X_{n+1}$, $P_{S_n} X_{n+2}$, ... in that order



$\underline{\text{Remarks}}$: • It may seem unreasonable to assume that $X_{1-d}, \dots, X_0 \perp Y_t$, $t \geq 1$. However, we can easily construct an example where this holds, for $d = 1$. Consider $X_t = X_0 + \sum_{j=1}^{t} Y_t$, $t \geq 1$, where $\{Y_t\}$ is ARMA$(p,q)$, for some RV $X_0 \perp Y_t$, $t \geq 1$. Then $\{X_t\}$ is an ARIMA$(p, 1, q)$ process.

- The prediction MSE can also be computed.

× $\underline{\text{Seasonal ARIMA (SARIMA)}}$.

In addition to removing trends, we apply the difference operator to remove the seasonal components

$\underline{\text{Ex}}$: Pure Seasonal ARMA.

Consider monthly observations, gathered over the course of $r$ years.

| Year \ Month | 1 | 2 | ... | 12 |
|---|---|---|---|---|
| 0 | $X_{-11}$ $U_{-11}$ | $X_{-10}$ $U_{-10}$ | | $X_0$ $U_0$ |
| 1 | $X_1$ $U_1$ | $X_2$ $U_2$ | | $X_{12}$ $U_{12}$ |
| 2 | $X_{13}$ $U_{13}$ | $X_{14}$ $U_{14}$ | | $X_{24}$ $U_{24}$ |
| 3 | $X_{25}$ $U_{25}$ | $X_{26}$ $U_{26}$ | | $X_{36}$ $U_{36}$ |
| ⋮ | | | | |
| r | $X_{1+12(r-1)}$ | $X_{2+12(r-1)}$ | | $X_{12+12(r-1)}$ |

To account for seasonal effects, suppose that each time series restricted to a particular month follows an ARMA$(P, Q)$ model.

(67)

$\Rightarrow$ For $j = 1, \dots, 12$, $t = 0, \dots, r-1$,

$$X_{j+12t} = \Phi_1 X_{j+12(t-1)} + \dots + \Phi_P X_{j+12(t-P)} + U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \dots + \Theta_Q U_{j+12(t-Q)},$$

where $$\begin{cases} U_1, U_{13}, \dots \sim WN(0, \sigma_u^2) \\ U_2, U_{14}, \dots \sim WN(0, \sigma_u^2) \\ \vdots \\ U_{12}, U_{24}, \dots \sim WN(0, \sigma_u^2) \end{cases}$$

For example, with $P = Q = 1$, expanding the difference equation,

$$\begin{aligned} X_1 &= \Phi_1 X_{-11} + \Theta_1 U_{-11} + U_1 \\ X_2 &= \Phi_1 X_{-10} + \Theta_1 U_{-10} + U_2 \\ &\vdots \\ X_{12} &= \Phi_1 X_0 + \Theta_1 U_0 + U_{12} \\ X_{13} &= \Phi_1 X_1 + \Theta_1 U_1 + U_{13}, \end{aligned}$$

and we see that for all $t$, $X_t = \Phi_1 X_{t-12} + \Theta_1 U_{t-12}$

More generally, $\forall t$,

$$X_t = \Phi_1 X_{t-12} + \dots + \Phi_P X_{t-12P} + U_t + \Theta_1 U_{t-12} + \dots + \Theta_Q U_{t-12Q}$$

$\Leftrightarrow$ $\Phi(B^{12}) X_t = \Theta(B^{12}) U_t$, where

$$\Phi(z) = 1 - \Phi_1 z - \dots - \Phi_P z^P$$
$$\Theta(z) = 1 + \Theta_1 z + \dots + \Theta_Q z^Q$$
$$\{U_{j+12t}, t \in \mathbb{Z}\} \sim WN(0, \sigma_u^2) \quad \forall j.$$

$\Phi(B^s) X_t = \Theta(B^s) U_t$ is called an $ARMA(P,Q)_s$ process, $s > 0$

Pure Seasonal

(68)

To model dependence between the series $\{U_{j+12t}\}$, we may assume that the sequence $\{U_t\}$ follows an $ARMA(p,q)$ model $\phi(B) U_t = \theta(B) Z_t$, $Z_t \sim WN(0, \sigma^2)$, yielding

$$\begin{aligned} \Phi(B^s) X_t &= \Theta(B^s) \phi^{-1}(B) \theta(B) Z_t \\ &= \phi^{-1}(B) \theta(B) \Theta(B^s) Z_t \end{aligned}$$

commutativity

i.e. $$\phi(B) \Phi(B^s) X_t = \theta(B) \Theta(B^s) Z_t.$$

More generally, we may detrend first the process $\{X_t\}$, leading us to the general $SARIMA(p,d,q) \times (P,D,Q)_s$ process:

Let $q, D \geq 0$

$\{X_t\}$ is said to be a $SARIMA(p,d,q) \times (P,D,Q)_s$ process if the differenced process $Y_t := (1-B)^d (1-B^s)^D X_t$ is a causal ARMA process

$$\phi(B) \Phi(B^s) Y_t = \theta(B) \Theta(B^s) Z_t, \quad Z_t \sim WN(0, \sigma^2)$$

where

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$
$$\Phi(z) = 1 - \Phi_1 z - \dots - \Phi_P z^P, \quad \Theta(z) = 1 + \Theta_1 z + \dots + \Theta_Q z^Q$$

## References

[1] P. J. Brockwell & R. A. Davis
Time Series : Theory and Methods, Springer (2006)

[2] P. J. Brockwell & R. A. Davis
Introduction to Time Series and Forecasting, Springer
Second Edition (2002)

[3] P. Bartlett
Lecture Notes : Introduction to Time Series (2010)
(Berkeley) ( Stat 153, Fall 2010 )