Ch. 2 : Autoregressive Model

Contents

1 AR(1) model **3**
1.1 AR(1) 4
1.2 ACF and ACVF of AR(1) 9
1.3 Summary 1 17

2 Causal Representation and Stationarity Condition **18**
2.1 Causal Representation of AR(1) 19
2.2 AR(p) 27
2.3 Summary 2 35

3 Parameter Estimation **37**
3.1 Yule-Walker Equation 38
3.2 Partial ACF 50
3.3 Summary 3 55

4 Fitting AR model **56**
4.1 De-mean or Not 57
4.2 Order Selection of AR(p) 60
4.3 Summary 4 70

5 Forecasting AR(p) 71
5.1 Best Linear Predictor 72
5.2 One-Step Prediction of AR(p) 79
5.3 Summary 5 89

6 Additional Topic - Linear Process 90
6.1 Linear Process 91
6.2 Calculating ACVF of AR(p) using Linear Process 95
6.3 Large sample property of sample ACF 103

February 17, 2017

1

AR(1) model

[ToC]

1.1 AR(1)

[ToC]

First order autoregression process is defined as

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

and ϕ is real-valued constant.

If $\phi = .8$, then

$$X_t = (.8)X_{t-1} + e_t$$

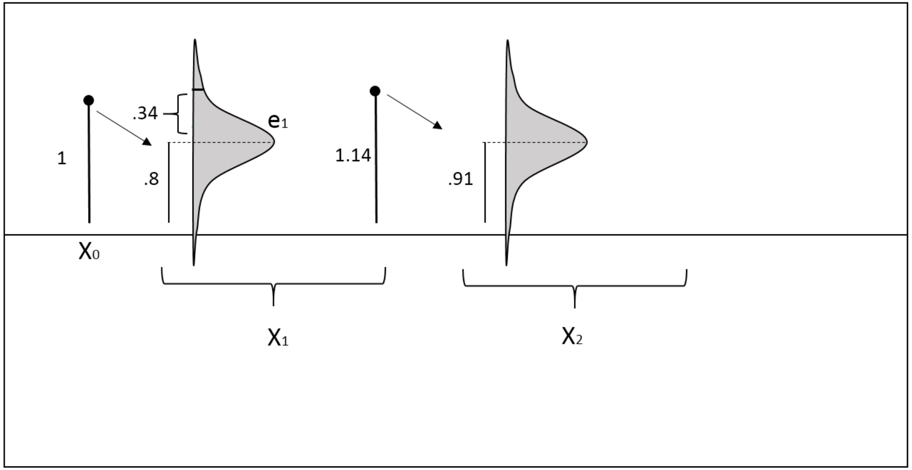
Auto-Regressive

When $\phi = .8$, then

Table 1:

t	0	1	2	3	4	5 $\cdots$
ϕX_{t-1}	-	.80	.91	1.26	0.28	$\cdots$
ϵ_t	-	.34	.66	-.91	-.35	$\cdots$
X_t	1	1.14	1.57	0.35	-.07	$\cdots$

Inside AR(1)



Questions:

- Is this series stationary? Does ϕ has anything to do with the stationarity?
- What is $E(X_t)$, $V(X_t)$ and $\gamma(h)$ of this process?

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

Stationarity Condition for AR(1)

- We need $|\phi| < 1$ for AR(1) process to be stationary.

Mean of AR(1)

- It will be shown in the next section (2.1) that

$$E(X_t) = 0$$

1.2 ACF and ACVF of AR(1)

- Assume that the AR(1) is stationary. (i.e. $|\phi| < 1$)
- What is the theoretical ACVF of AR(1)? Let's start with variance ($h = 0$).

$$\begin{aligned}
\gamma(0) &= \mathrm{Var}(X_t) \quad = \quad \mathrm{Cov}(X_t, X_t) \\
&= \mathrm{Var}(\phi X_{t-1} + e_t) \qquad\qquad \text{(because } X_{t-1} \text{ and } e_t \text{ are independent,)} \\
&= \mathrm{Var}(\phi X_{t-1}) + \mathrm{Var}(e_t) \\
&= \phi^2 \mathrm{Var}(X_{t-1}) + \sigma^2 \quad = \quad \phi^2\,\gamma(0) + \sigma^2
\end{aligned}$$

- We have

$$\gamma(0) = \phi^2\,\gamma(0) + \sigma^2.$$

- Given

$$\gamma(0) \quad = \quad \phi^2\,\gamma(0) + \sigma^2,$$

solve for $\gamma(0)$, we get formula for variance of X_t,

$$\gamma(0) = \sigma^2/(1 - \phi^2).$$

- Do you see what's wrong if we have $|\phi| > 1$?

When $h = 1$

Still assuming the stationarity, let's look at when h is not 0,

$$\begin{aligned}
\gamma(1) &= \mathrm{Cov}(X_{t+1}, X_t) \\
&= \mathrm{Cov}\big(\phi\, X_t + e_{t+1}, \quad X_t\big) \\
&= \mathrm{Cov}\big(\phi\, X_t,\ X_t\big) + \mathrm{Cov}\big(e_{t+1},\ X_t\big) \\
&= \phi\, \mathrm{Cov}(X_t, X_t) \quad = \quad \phi\, \gamma(0)
\end{aligned}$$

When $h = 2$

$$
\begin{aligned}
\gamma(2) &= \mathrm{Cov}(X_{t+2}, X_t) \\
&= \mathrm{Cov}\big(\phi\, X_{t+1} + e_{t+2}, \quad X_t\big) \\
&= \mathrm{Cov}\big(\phi\, X_{t+1},\ X_t\big) + \mathrm{Cov}\big(e_{t+2},\ X_t\big) \\
&= \phi\, \mathrm{Cov}(X_{t+1}, X_t) \quad = \quad \phi\, \gamma(1) \quad = \quad \phi^2 \gamma(0)
\end{aligned}
$$

ACF and ACVF of AR(1)

To Summary

So the ACVF of AR(1) looks like

$$\gamma(h) = \begin{cases} \frac{\sigma^2}{(1-\phi^2)} & \text{for } h = 0 \\ \\ \phi^{|h|}\,\gamma(0) & \text{for } h > 0 \end{cases}$$

Then since ACF $=$ ACVF $/\gamma(0)$,

$$\rho(h) = \begin{cases} 1 & \text{for } h = 0 \\ \phi^{|h|} & \text{for } h > 0 \end{cases}$$

Alternative Notation

- We could also write AR(1) as

$$X_t - \phi X_{t-1} = e_t,$$

and using the **backward operator**, write

$$\underbrace{(1 - \phi B)}_{\Phi(B)} X_t = e_t.$$

$\Phi(z)$ is called **characteristic polynomial** of AR(1).

- Backward operator makes it go back a day:

$$BX_t = X_{t-1}.$$

ACVF of AR(1)

Let $\phi = .8.$ i.e. $\quad \Phi(z) = 1 - .8z$

To Summary

```
install.packages("ltsa")  #- if not installed before

library(ltsa)

ACVF  <- tacvfARMA(phi = c(.8), theta= c(0), maxLag=20, sigma2=1)   #- get Theoretical ACVF
ACF   <- ACVF/ACVF[1]                                               #- get Theo. ACF


plot(0:20, ACVF, type="h", col="red");   abline(h=0)
plot(0:20, ACF,  type="h", col="red");   abline(h=0)


#-  you can also use (ACF only)
ACF2  <- ARMAacf(ar=c(.8), ma=c(0), lag.max=20, pacf=FALSE )
PACF2 <- ARMAacf(ar = Fit3$ar, ma = numeric(), lag.max = 20, pacf = TRUE)
```

Simulating of AR(1)

Let $\phi = .8$. i.e. $\quad \Phi(z) = 1 - .8z$

To Summary

```
Y <- arima.sim(list(ar = c(.8 ) ), 100 )     #- Simulate AR(1) with phi=.8
plot(Y, type="o")

ACVF1  <- tacvfARMA(phi = c(.8), maxLag=20)  #- need library(ltsa)
ACF1   <- ACVF1/ACVF1[1]                     #- get Theo. ACF

acf(Y)                             #- sample ACF
lines(0:20, ARrho, col="red")      #- overlay theoretical ACF
```

1.3 Summary 1

[ToC]

---

1. AR(1) process is defiend by formula

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

2. Which is alternative written using characteristic polynomial and backwards operator,

$$(1 - \phi B)X_t = e_t$$

3. $|\phi|$ needs to be less than 1, for AR(1) process to be stationary.
4. Theoretical ACF and ACVF of AR(1) are listed on p.13
5. You can compute numerical value of above formula in R as on p.15
6. You can simulate AR(1) process as on p.16

2

Causal Representation and AR(p) Process

[ToC]

2.1 Causal Representation of AR(1)

[ToC]

- If AR(1) representation is

$$X_t = \phi X_{t-1} + e_t,$$

then I can write the same thing for yesterday,

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

- Starting from usual AR(1) representation,

$$\begin{aligned} X_t &= \phi\, X_{t-1} + e_t \\ &= \phi\Big\{\phi X_{t-2} + e_{t-1}\Big\} + e_t \\ &= \phi^2 X_{t-2} \quad + \quad \phi\, e_{t-1} + e_t \end{aligned}$$

Do it again , and we get

$$\begin{aligned} X_t &= \phi^2 X_{t-2} \quad + \quad \phi\, e_{t-1} + e_t \\ &= \phi^2 \big(\phi X_{t-3} + e_{t-2}\big) \quad + \quad \phi\, e_{t-1} + e_t \\ &= \phi^3 Y_{t-3} \quad + \quad \phi^2\, e_{t-2} + \phi e_{t-1} + e_t \end{aligned}$$

We can keep doing this, and get

$$X_t = \phi^k X_{t-k} \quad + \quad e_t + \phi\, e_{t-1} \cdots \phi^{k-1} e_{t-(k-1)}$$

If $|\phi| < 1$, then letting $k \to \infty$ will yield

$$X_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots \quad = \sum_{i=0}^{\infty} \phi^i\, e_{t-i}$$

Causal Representation

- So here's another ways to write AR(1) process

$$
\begin{aligned}
X_t &= \sum_{i=0}^{\infty} \phi^i \, e_{t-i} \\
&= e_t + \phi \, e_{t-1} + \phi^2 \, e_{t-2} + \cdots
\end{aligned}
$$

- This is called **causal representation**, because we can write Y_t as infinite sum of **past** errors (innovations).

- Now it is easy to see that mean of AR(1)

$$
E(X_t) = \sum_{i=0}^{\infty} \phi^i \, E(e_{t-i}) \quad = \quad 0
$$

- Recall e_t are White Noise with mean 0 and variance σ.

- Or often, we assume

$$
e_t \sim N(0, \sigma^2)
$$

Causal Representation and Characteristic Polynomial

- Recall yet another way to write AR(1)

$$(1 - \phi B)\, X_t \;=\; e_t.$$

This means that we can write

$$X_t \;=\; \frac{1}{(1 - \phi\, B)}\, e_t.$$

- Compare this with Causal representation

$$\begin{aligned} X_t &= e_t + \phi\, e_{t-1} + \phi^2\, e_{t-2} + \cdots \\ &= (1 + \phi B + \phi^2 B^2 + \cdots)\, e_t \end{aligned}$$

- So we have the equivalence on the right hand side,

$$\frac{1}{(1-\phi B)} \quad = \quad 1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \cdots$$

- This is exactly same as the geometric series,

$$\frac{1}{1-x} \quad = \quad 1 + x + x^2 + x^3 + \cdots$$

- Note that the condition for the geometiric series is

$$|x| < 1 \qquad \text{or} \qquad |\phi| < 1$$

  Which is same as the **causal (stationary) condition**.

Another look at Causal Condition

- So if you write AR(1) in causal way,

$$X_t = \sum_{i=0}^{\infty} \phi^i \, e_{t-i},$$

- Then we can calculate variance as

$$\begin{aligned} \mathrm{Var}(X_t) &= \mathrm{Var}\left[\sum_{i=0}^{\infty} \phi^i \, e_{t-i}\right] &= \sum_{i=0}^{\infty} \mathrm{Var}(\phi^i \, e_{t-i}) \\ &= \sum_{i=0}^{\infty} \phi^{2i} \mathrm{Var}(e_{t-i}) &= \sigma^2 \sum_{i=0}^{\infty} \phi^{2i} \end{aligned}$$

  This does not converge unless $|\phi| < 1$.

- When it does converge, the variance is $\sigma^2/(1-\phi^2)$.

Causal Condition of AR(1)

We can represent AR(1) in causal representation only when $|\phi| < 1$.

$$X_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots$$

- When AR(1) has $|\phi| > 1$, and if you represent that with past errors, then it is called explosive process, and it's not stationary.
- When AR(1) has $|\phi| > 1$, Y_t can be written as infinite sum of future errors, and it is a unique stationary solution to AR(1) equation.
- When $|\phi| = 1$, then there is no stationary solution. What is the other name of Y_t when $\phi = 1$?
- We will assume that all AR process we deal with are causal.
- If AR process is not causal, then it can be re-written as causal process with different innovations. (Prob 3.8).

Simulating non-causal AR(1)

Let $\phi = 1.2$. i.e. $\Phi(z) = 1 - 1.2z$

```
Y <- arima.sim(list(ar = c(1.2) ), 100 )      #- gives error



#- Hand written simulation
Y <- 0.5
phi <- 1.2
e <- rnorm(100)
for (t in 2:100){
  Y[t] = Y[t-1]*phi + e[t]
}
plot(Y, type="o")
```

2.2 AR(p)

[ToC]

- Autoregressive process of order p is

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} \;=\; e_t,$$

where $e_t \sim WN(0, \sigma^2)$

and $\phi_1, \ldots, \phi_p$ is real valued constant.

- Alternative notation using characteristic polynomial is,

$$\underbrace{\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)}_{\Phi(B)} X_t \;=\; e_t.$$

$$\Phi(B) \;\; X_t \;=\; e_t.$$

Writing AR(p) in Causal Rep.

- Using the characteristic polynomial,

$$
\begin{aligned}
\Phi(B)\, X_t &= e_t \\
X_t &= \frac{1}{\Phi(B)}\, e_t
\end{aligned}
$$

- So if we could write AR(p) as causal,

$$
\begin{aligned}
X_t &= \psi_0 + \psi_1 e_t + \psi_2 e_{t-2} + \psi_3 e_{t-3} + \cdots \\
&= \left(\psi_0 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots\right) e_t
\end{aligned}
$$

- That means we must be able to write polynomial as

$$
\frac{1}{\left(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p\right)} = \psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \cdots
$$

- When can we do this?

Causal Condition

- From operator theory, we know that if

(complex) root of $\Phi(z)$ is outside of the unit circle,

we can expand the inverse of $\Phi(z)$ as

$$\frac{1}{\Phi(z)} = \psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \cdots$$

- This is the condition that allows us to write AR(p) in causal representation.

Causal Conditon in AR(p)

- We have seen that causal condition for AR(1) was $|\phi| < 1$.
- This was because polynomial

$$\Phi(z) \quad = \quad (1 - \phi z)$$

will have root inside the unit circle if $|\phi| > 1$.

- AR(p) will admit the causal representation if the characteristic polynomial

$$\Phi(z) = 1 + \phi_1 z + \phi_2 z^2 + \phi_3 z^3 + \cdots + \phi_p z^p$$

has all the roots outside of the unit circle. Causal representation will ensure stationarity.

Example: Checking Causality 1

Check to see of AR(2) model,

$$Y_t = .4Y_{t-1} - .3Y_{t-2} + e_t$$

is causal (stationary).

We have to look at the root of

$$\Phi(z) \quad = \quad 1 - .4z + .3z^2$$

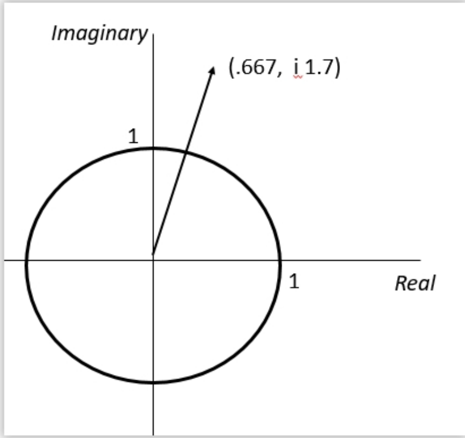
which is

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{.4 \pm \sqrt{(.4)^2 - 4(.3)(1)}}{2(.3)} = .667 \pm i1.7$$

Their distance form the origin is

$$\sqrt{(.667)^2 + (i1.7)^2} = 1.826$$

So the roots are outside the complex unit circle. Thus this AR(2) is causal.



Example: Checking Causality 2

Check to see of AR(2) model,

$$Y_t = -.7Y_{t-1} - .6Y_{t-2} + e_t$$

is causal or not.

We have to look at the root of

$$\Phi(z) = 1 + .7z + .6z^2,$$

which has roots $-.583 \pm 1.15i$. They are at at distance $\sqrt{(-.583)^2 + (1.15i)^2} = 1.29$ from origin. Therefore, this AR(2) is causal.

```
z <- polyroot(c(1,.7,.6))   # find root of 1+.7z+.6z^2
z                           #- see the root
Mod(z)                      #- Distance from origin
```

Example: Checking Causality 3

Given

$$X_t = .7X_{t-1} + .6X_{t-2} + e_t,$$

Check the causality.

we look at the root of

$$\Phi(z) = 1 - .7z - .6z^2,$$

which has roots $.833, .2$. Therefore, this AR(2) is not causal.

```
z <- polyroot(c(1,-.7,-.6))
z
Mod(z)
```

2.3 Summary 2

[ToC]

---

1. AR(p) is defined as

$$\begin{aligned} X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} &= e_t, \\ \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) X_t &= e_t. \\ \Phi(B)\ X_t &= e_t. \end{aligned}$$

   where $\phi_1, \ldots, \phi_p$ is real valued constant, and $e_t \sim WN(0, \sigma^2)$.

2. AR(p) can be written in causal representation,

$$X_t = \sum_{i=0}^{\infty} \phi^i \, e_{t-i},$$

   when its characteristic polynomial $\Phi(z)$ has all the roots **outside** of the unit circle on the imaginary plane.

3. When the AR(p) can be written as causal process, then it is stationary.

4. You can use `polyroot()` function in R to calculate the roots of polynomial, and `Mod()` to calculate their distance from the origin.

```
z <- polyroot(c(1,.7,.6))   # find root of 1+.7z+.6z^2
z                           #- see the root
Mod(z)                      #- Distance from origin
```

Parameter Estimation in AR(p)

[ToC]

3.1 Yule-Walker Equation

[ToC]

We start with AR(p) equation. Let $p = 3$ for now,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \phi_3 X_{t-3} \quad = \quad e_t.$$

We multiply both sides by X_t, and take expectation.

$$E\Big(X_t \left[X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \phi_3 X_{t-3}\right]\Big) \quad = \quad E\Big(X_t \left[e_t\right]\Big).$$

Recall the formula for covariance,

$$\mathrm{Cov}(X_t, X_{t-1}) = E(X_t\, X_{t-1}) - E(X_t)E(X_{t-1}).$$

So if $E(X_t) = 0$, then

$$\gamma(1) = E(X_t\, X_{t-1}).$$

Then the equation in the last page

$$E\Big(X_t\,[X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \phi_3 X_{t-3}]\Big) \;=\; E\Big(X_t\,[e_t]\Big).$$

can be written as

$$\gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2) - \phi_3\gamma(3) = \sigma^2$$

Recall $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + e_t$

If we use the original AR(3) equation for Y_{t+1} instead of for Y_t,

$$X_{t+1} - \phi_1 X_t - \phi_2 X_{t-1} - \phi_3 X_{t-2} = e_{t+1}$$

then we would have gotten

$$E\Big(X_t \left[X_{t+1} - \phi_1 X_t - \phi_2 X_{t-1} - \phi_3 X_{t-2}\right]\Big) = E\Big(X_t \left[e_{t+1}\right]\Big).$$

$$\gamma(1) - \phi_1\gamma(0) - \phi_2\gamma(1) - \phi_3\gamma(2) = 0.$$

If we use the original AR(3) equation for X_{t+2} instead of for X_t, we would get

$$E\Big(X_t\left[X_{t+2} - \phi_1 X_{t+1} - \phi_2 X_t - \phi_3 X_{t-1}\right]\Big) \quad = \quad E\Big(X_t\left[e_{t+1}\right]\Big).$$

$$\gamma(2) - \phi_1\gamma(1) - \phi_2\gamma(0) - \phi_3\gamma(1) \quad = \quad 0.$$

Repeat it one more time, and we get equations,

$$\gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2) - \phi_3\gamma(3) \quad = \quad \sigma^2$$

$$\begin{aligned}
\gamma(1) - \phi_1\gamma(0) - \phi_2\gamma(1) - \phi_3\gamma(2) &= 0\\
\gamma(2) - \phi_1\gamma(1) - \phi_2\gamma(0) - \phi_3\gamma(1) &= 0\\
\gamma(3) - \phi_1\gamma(2) - \phi_2\gamma(1) - \phi_3\gamma(0) &= 0.
\end{aligned}$$

We'll keep the first equation, and re-write the rest of equations

$$
\begin{aligned}
\gamma(1) - \phi_1\gamma(0) - \phi_2\gamma(1) - \phi_3\gamma(2) &= 0 \\
\gamma(2) - \phi_1\gamma(1) - \phi_2\gamma(0) - \phi_3\gamma(1) &= 0 \\
\gamma(3) - \phi_1\gamma(2) - \phi_2\gamma(1) - \phi_3\gamma(0) &= 0
\end{aligned}
$$

as

$$
\begin{aligned}
\phi_1\gamma(0) + \phi_2\gamma(1) + \phi_3\gamma(2) &= \gamma(1) \\
\phi_1\gamma(1) + \phi_2\gamma(0) + \phi_3\gamma(1) &= \gamma(2) \\
\phi_1\gamma(2) + \phi_2\gamma(1) + \phi_3\gamma(0) &= \gamma(3).
\end{aligned}
$$

Which can be put in a matrix form as

$$
\begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix}
\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \end{bmatrix}.
$$

We still have the first equation,

$$\gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2) - \phi_3\gamma(3) \quad = \quad \sigma^2.$$

These two equations are called **Yule-Walker Equations**.

Yule-Walker Equations

For AP(3),

$$\gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2) - \phi_3\gamma(3) \quad = \quad \sigma^2$$

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \end{bmatrix}.$$

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \end{bmatrix}.$$

Yule-Walker Estimators for AR(p)

To Summary

We can use Y-W equasion backwards with sample ACVF to estimate parameters.

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \end{bmatrix} = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \hat{\gamma}(1) \\ \hat{\gamma}(2) & \hat{\gamma}(1) & \hat{\gamma}(0) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\gamma}(1) \\ \hat{\gamma}(2) \\ \hat{\gamma}(3) \end{bmatrix}.$$

$$\hat{\boldsymbol{\phi}}_{\mathbf{3}} = \hat{\boldsymbol{\Gamma}}_{\mathbf{3}}^{\mathbf{-1}} \, \hat{\boldsymbol{\gamma}}_{\mathbf{3}}$$

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}_1 \hat{\gamma}(1) - \hat{\phi}_2 \hat{\gamma}(2) - \hat{\phi}_3 \hat{\gamma}(3)$$

Example: Y-W estimator

```
Y <- arima.sim(list(ar = c(.6, .3) ), 100 )

layout(matrix(1:3, 3,1))  #- 3 plot at once
plot(Y, type="o")
acf(Y)
pacf(Y)
layout(1)                 #- reset the layout

Fit1 <- ar(Y, aic=F, order.max=2)                #- fits AR(2) by Y-W(default)
```

Large sample property of Yule-Walker Estimator

To Summary

$\sqrt{n}(\hat{\phi}_p - \phi_p)$ is approximately $\mathcal{N}(0, n^{-1}\sigma^2\boldsymbol{\Gamma}^{-1})$ if n is large.

95% confidence interval for ϕ_j is

$$\hat{\phi}_j \pm 1.96\sqrt{\frac{\hat{\sigma}^2\,\hat{\Gamma}_{jj}^{-1}}{n}}$$

where $\hat{\Gamma}_{jj}^{-1}$ be jth diagnal element of $\hat{\boldsymbol{\Gamma}}^{\mathbf{-1}}$,

If you use `ar()` function, `$asy.var.coef` is same as $\sigma^2\boldsymbol{\Gamma}^{-1}/n$

Example: Testing AR parameters for significance

```
Y     <- arima.sim(list(ar = c(.6, .3) ), 100 ) #- Simulate AR(2)

Fit1 <- ar(Y, aic=F, order.max=2)        #- fits AR(2) by Y-W(default)

str(Fit1)                                #- see what’s inside

phi   <- Fit1$ar                         #- phi1 hat and phi2 hat
sigSq <- Fit1$var.pred                   #- sig^2 hat
phiSE <- sqrt(Fit1$asy.var.coef[1,1])    #- standard error for phi hats

Fit1$asy.var.coef

phi                              #- check if phi hats are significant
phi1SE
```

```
#- If you want to calculare asy.var.coef by hand -
  A <- acf(Y, type="covariance")$acf
  G <- matrix(A[c(1,2,2,1)], 2,2)
  sigSq * solve(G) / 100                  #- compare to asy.var.coef
```

3.2 Partial ACF

[ToC]

Recall the Yule-Walker Equation,

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \end{bmatrix}.$$

$$\boldsymbol{\phi_3} = \boldsymbol{\Gamma_3^{-1}} \, \boldsymbol{\gamma_3}$$

Note that this equation, can be extended to more than 3 parameters, even though we only have 3 ϕ's.

For example, if we use $\mathbf{\Gamma}_5$, then

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \gamma(4) \\ \gamma(1) & \gamma(0) & \gamma(1) & \gamma(0) & \gamma(3) \\ \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(4) & \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \gamma(4) \\ \gamma(5) \end{bmatrix}$$

$$\boldsymbol{\phi}_5 = \mathbf{\Gamma}_5^{-1} \, \boldsymbol{\gamma}_5 .$$

PACF

Partial ACF of lag h is defined as last element of vector

$$\boldsymbol{\phi_h} \;=\; \boldsymbol{\Gamma_h^{-1}}\;\boldsymbol{\gamma_h}.$$

For AR(p),

$$\begin{cases} \alpha(0) = 1 & \\ \alpha(h) = \phi_h & \text{if } 1 < h \leq p \\ \alpha(h) = 0 & \text{if } p < h \end{cases}$$

PACF of AR(p) cuts off after lag p.

Sample PACF

Sample version of PACF of lag h is the last element of

$$\hat{\boldsymbol{\phi}}_{\boldsymbol{h}} \quad = \quad \hat{\boldsymbol{\Gamma}}_{\boldsymbol{h}}^{-1} \; \hat{\boldsymbol{\gamma}}_{\boldsymbol{h}}.$$

For AR(p),

$$\begin{cases} \alpha(0) = 1 & \\ \alpha(h) = \phi_h & \text{if } 1 < h \leq p \\ \alpha(h) = 0 & \text{if } p < h \end{cases}$$

PACF of AR(p) cuts off after lag p.

Example: ACF and PACF of AR(p)

```
X <- arima.sim(list(ar = c(.6, .3) ), 100 ) #- Simulate AR(2)
```

```
plot(X)
acf(X)
pacf(X)
```

```
layout(matrix(1:3, 1,3))    #- Plot them at once
plot(X, type="o")
acf(X);
pacf(X)
layout(1)                   #- reset layout
```

3.3 Summary 3

[ToC]

---

1. One method for estimating parameters in AR(p) is to use Yule-Walker Estimator.
2. Yule-Walker Estimator make use of relationship between ACVF and AR(p) parameters.
3. When n is large, standard error of Y-W estimator can be calculated using large-sample formula.
4. Partial ACF (PACF) is defined on p. 48, and it is as characteristc of AR(p) process as ACF.
5. ACF for AR(p) process decays, as PACF for AR(p) cuts off after lag p.
6. You can use `pacf()` function in R to plot PACF.

4

Fitting AR model

[ToC]

4.1 De-mean or Not

[ToC]

- Theoretically, AR(p) has mean of 0.
- De-mean is when we have

$$Y_t = \mu + X_t, \qquad \text{where } \mu \text{ is a constant and } X_t \text{ is AR(p)},$$

then $E(Y_t) = \mu$. We need to de-trend by letting

$$\hat{X}_t = Y_t - \bar{Y}.$$

and model $\hat{X}_t$ with zero-mean AR(p).

- `detrend=TRUE` is default in `ar()`

- `detrend=FALSE` option.
  Our model is

$$X_t = Y_t.$$

  So $E(X_t) = 0$. we directly model the data with zero-mean AR(p) model.

- So if we let

```
Y2 = Y-mean(Y)
```

  Then use `ar(Y2, demean=FALSE)`, then it is same as using `demean=TRUE` option.

Example: De-mean or not

```
Y    <- arima.sim(list(ar = c(.6, .3) ), 100 )  + 8.5
plot(Y, type="o")


Fit1 <- ar(Y, aic=F, order.max=2)       #- demean=TRUE is the default
Fit1

Fit2 <- ar(Y, aic=F, order.max=2, demean=T)
Fit2

Fit3 <- ar(Y, aic=F, order.max=2, demean=F)
Fit3

Fit4 <- ar(Y-mean(Y), aic=F, order.max=2, demean=F)
Fit4
```

4.2 Order Selection of AR(p)

[ToC]

---

1. ACF and PACF
2. AIC
3. Parameter significance

1. ACF and PACF

- ACF of AR(p) tails off
- PACF of AR(P) cuts off at lag p.

2. AIC

Akaike Information Criteria:

$$\mathrm{AIC} = -2\ \log(\text{maximum likelihood}) + 2k$$

$k = p + 1$ if `demean=TRUE` and $k = p$ if `demean=FALSE`.

Choose p that **MINIMIZE** AIC.

3. Parameter Significance

If true model is AR(2),

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

but if you fit AR(3), then

$$\begin{aligned} E(\hat{\phi}_1) &= \phi_1 \\ E(\hat{\phi}_2) &= \phi_2 \\ E(\hat{\phi}_3) &= 0 \end{aligned}$$

Test the pameter estimate for significance using CI.

Example

```
Y    <- arima.sim(list(ar = c(.6, .3) ), 100 )   #- simulate AR(2)


#- 1. ACF and PACF
layout(matrix(1:2, 1,2))
acf(Y); pacf(Y)
layout(1)


#- 2. AIC
Fit1 <- ar(Y)                            #- Fit AR(p).  find best p by AIC
Fit1


#- 3. Parameter Significance
Fit1 <- ar(Y, aic=F, order.max=3 )       #- fit AR(p). force p=3.

Fit1$ar                                  #- phi1 hat and phi2 hat
1.96*sqrt(Fit1$asy.var.coef)             #- standard error for phi hats
```

Residual of AR(p)

After parameter estimation, say we have AR(2) model

$$Y_t = \hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2} + e_t.$$

Y_t are observations. But we never get to observe e_t.

We let "residuals" to be

$$\hat{e}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} \qquad t = 3, 4, \ldots, n$$

Example

```
Y <- arima.sim(list(ar = c(.6, .3) ), 100 )

Fit1 <- ar(Y, aic=F, order.max=2, demean=F)
Fit1$ar

Y[100]-Fit1$ar[1]*Y[99]-Fit1$ar[2]*Y[98]

Fit1$resid
```

Simulation Study

If we use AIC all the time, what is the probability we end up with correct p?

- Set $p = 2$. Generate simulated AR(2).
- Use `ar(Y, order.max=5)` and search for best p.
- Repeat 1000 times. How many time do we end up with right p?

```
#- Simulation Study of choosing p by AIC

    Result <- 0
    for (i in 1:1000){

      Y <- arima.sim(list(ar = c(.6, .3) ), 100 )
      Fit1 <- ar(Y, aic=T, order.max=5)

      print(c(i, Fit1$order) )     #- print order on screen (optional)

      Result[i] <- (Fit1$order==2)

    }
    Result

    sum(Result)/1000
```

Example: Dow Jones 1972

```
D  <- read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/dowj.csv")
D1 <- ts(D, start=c(1,1), freq=1)
Y <- diff(log(D1))

plot(Y, type='o')
acf(Y)
pacf(Y)

Fit1 <- ar(Y)                              #- Fit AR(p). find best p by AIC.  Use Y-W estimator, demean=TRUE
Fit2 <- ar(Y, aic=TRUE,  order.max=20)     #- best p is chosen by aic (same as above)
Fit3 <- ar(Y, aic=FALSE, order.max=5)      #- Fit AR(p). force p=5. (do not choose p by AIC)

Fit1                                  #- see the summary of fit

Fit1$aic                              #- look at aic

Fit1$ar                               #- Y-W parameter estimate
1.96*sqrt( diag(Fit31asy) )           #- 1.95*(SE of Y-W) for checking parameter significance

e.hat <- Fit1$resid                        #- extract residuals after the fit
plot(e.hat, type="o", na.action=na.pass)   #- plot residuals

acf(e.hat)                            #- check for autocorrelation in residuals
pacf(e.hat)
```

4.3 Summary 4

[ToC]

---

1. To fit AR(p), you must first decide to de-mean or not. `ar()` default is `demean=TRUE`. It is more customary to just de-mean.

2. Look at sample ACF and sample PACF to guess p.

3. Akaike Information Criteria can also be used. `ar()` function uses it by default. You want MINIMUM AIC.

4. Check parameter significance of estimated parameters.

5. Check residuals for correlation.

6. Sample code is on page 70

5

Forecasting AR(p)

[ToC]

5.1 Best Linear Predictor

[ToC]

- Given data $X_1, \ldots, X_n$, we want to predict X_{n+1}.

- Let $\hat{X}(1)$ be 1-step predictor given $X_1, \ldots, X_n$.

  1. We only consider linear predictor

  $$\hat{X}(1) = a_0 + a_1 X_n + \cdots + a_n X_1$$

  2. We define 'best predictor' as predictor that has minumim prediction MSE.

  $$\text{MSE} = E\Big[\big(\hat{X}(1) - X_{n+1}\big)^2\Big].$$

- Let $n = 4$ for simplicity. Then $h-$step linear predictor is

$$\hat{X}(h) = a_0 + a_1 X_4 + a_2 X_3 + a_3 X_2 + a_4 X_1$$

We are trying to find coefficients $\{a_0, a_1, \ldots, a_n\}$ that minimize MSE,

$$E\Big[(\hat{X}(h) - X_{n+h})^2\Big] = E\Big[\big(a_0 + a_1 X_n + \cdots + a_n X_1 - X_{n+h}\big)^2\Big].$$

- We'll take $\frac{\partial}{\partial a_i}$, and set them equal to 0.

- First, we take $\frac{\partial}{\partial a_0}$. Letting it go inside the expectation,

$$\begin{aligned}
\frac{\partial}{\partial a_0} E\Big[(\hat{X}(h) - X_{n+h})^2\Big] &= \frac{\partial}{\partial a_0} E\Big[(a_0 + a_1 X_4 + a_2 X_3 + a_3 X_2 + a_4 X_1 \quad - \quad X_{4+h})^2\Big] \\
&= E\Big[2(a_0 + a_1 X_4 + a_2 X_3 + a_3 X_2 + a_4 X_1 \quad - \quad X_{4+h})\Big] \\
&= 2a_0 + 2a_1 E\big[X_4\big] + 2a_2 E\big[X_3\big] + 2a_3 E\big[X_2\big] + 2a_4 E\big[X_1\big] \quad - \quad E\big[X_{4+h}\big] \\
&= 0
\end{aligned}$$

- That means, since $E(X_t) = 0$,

$$2a_0 \quad = \quad 0 \qquad \text{or,} \qquad a_0 = 0.$$

- Omitting $a_0 = 0$, now we take $\frac{\partial}{\partial a_1}$. We have

$$\begin{aligned}\frac{\partial}{\partial a_1} E\Big[(\hat{X}(h) - X_{n+h})^2\Big] &= \frac{\partial}{\partial a_1} E\Big[(a_1 X_4 + a_2 X_3 + a_3 X_2 + a_4 X_1 \quad - \quad X_{4+h})^2\Big] \\ &= E\Big[2(a_1 X_4 + a_2 X_3 + a_3 X_2 + a_4 X_1 \quad - \quad X_{4+h})\ X_4\Big] = 0\end{aligned}$$

- We can rewrite this, and get

$$a_1\gamma(0) + a_2\gamma(1) + a_3\gamma(2) + a_4\gamma(3) - \gamma(h) = 0,$$

or

$$a_1\gamma(0) + a_2\gamma(1) + a_3\gamma(2) + a_4\gamma(3) = \gamma(h).$$

- Now take $\frac{\partial}{\partial a_2}$. We have

$$\frac{\partial}{\partial a_2} E\Big[(\hat{X}(h) - X_{n+h})^2\Big] = E\Big[2(a_1 X_4 + a_2 X_3 + a_3 X_2 + a_4 X_1 \quad - \quad X_{4+h})\ X_3\Big] = 0.$$

And we get

$$a_1\gamma(1) + a_2\gamma(0) + a_3\gamma(1) + a_4\gamma(2) = \gamma(h+1).$$

- Now take $\frac{\partial}{\partial a_3}$. We have

$$\frac{\partial}{\partial a_3} E\Big[(\hat{Y}(h) - Y_{n+h})^2\Big] = E\Big[2(a_1 Y_n + \cdots + a_n Y_1 \quad - \quad Y_{n+h})\, Y_{n-2}\Big] = 0$$

We get

$$a_1\gamma(2) + a_2\gamma(1) + a_3\gamma(0) + a_4\gamma(1) = \gamma(h+2).$$

- Thus, we get set of equations,

$$\begin{aligned}
a_1\gamma(0) + a_2\gamma(1) + a_3\gamma(2) + a_4\gamma(3) &= \gamma(h) \\
a_1\gamma(1) + a_2\gamma(0) + a_3\gamma(1) + a_4\gamma(2) &= \gamma(h+1) \\
a_1\gamma(2) + a_2\gamma(1) + a_3\gamma(0) + a_4\gamma(1) &= \gamma(h+2) \\
a_1\gamma(3) + a_2\gamma(2) + a_3\gamma(1) + a_4\gamma(0) &= \gamma(h+3)
\end{aligned}$$

- This looks very similar to Yule-Walker equation,

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) \\ \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} \gamma(h) \\ \gamma(h+1) \\ \gamma(h+2) \\ \gamma(h+3) \end{bmatrix}.$$

- If $h = 1$, we have Yule-Walker equation, and

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) \\ \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \gamma(4) \end{bmatrix}.$$

  Therefore,

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{bmatrix}.$$

5.2 One-Step Prediction of AR(p)

- So the best 1-step ahead predictor for X_5 given $X_1, \ldots, X_4$ is

$$\hat{X}(1) = \phi_1 X_4 + \phi_2 X_3 + \phi_3 X_2 + \phi_4 X_1$$

- Note that our original AR(4) equation says,

$$X_5 = \phi_1 X_4 + \phi_2 X_3 + \phi_3 X_2 + \phi_4 X_1 + e_t$$

- In practice, we don't know the actual $\phi_1, \ldots, \phi_4$, so we must use the estimated version,

$$\hat{X}(1) = \hat{\phi}_1 X_4 + \hat{\phi}_2 X_3 + \hat{\phi}_3 X_2 + \hat{\phi}_4 X_1$$

h-step Prediction of AR(p)

- If $h = 2$, we have

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) \\ \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(2) \\ \gamma(3) \\ \gamma(4) \\ \gamma(5) \end{bmatrix}.$$

  Therefore,

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \boldsymbol{\Gamma}^{-\mathbf{1}}\boldsymbol{\gamma} \approx \hat{\boldsymbol{\Gamma}}^{-\mathbf{1}}\hat{\boldsymbol{\gamma}}$$

Prediction Error

- Minimized MSE is

$$E\Big[(\hat{X}(h) - X_{n+h})^2\Big] = E\Big[\big(\phi_1 X_4 + \phi_2 X_3 + \phi_3 X_2 + \phi_4 X_1 \quad - \quad X_{4+h}\big)^2\Big].$$

- If $h = 1$,

$$\begin{aligned} & E\Big[(\hat{X}(1) - X_5)^2\Big] \\ & \quad = \quad E\Big[\big(\phi_1 X_4 + \phi_2 X_3 + \phi_3 X_2 + \phi_4 X_1 \quad - \quad X_5\big)^2\Big] \\ & \quad = \quad E\Big[\Big\{\big(\phi_1 X_4 + \phi_2 X_3 + \phi_3 X_2 + \phi_4 X_1\big) \quad - \quad \big(\phi_1 X_4 + \phi_2 X_3 + \phi_3 X_2 + \phi_4 X_1 + e_t\big)\Big\}^2\Big] \\ & \quad = \quad E\Big[\ e_t^2\ \Big] \quad = \quad \sigma^2. \end{aligned}$$

- In practice, we don't know the actual σ^2, so we settle for

$$\hat{\text{MSE}} = \hat{\sigma}^2.$$

Large sample property of AR(p) predictor

Mean Error is zero

$$E\left[(\hat{X}(h) - X_{n+h})\right] = 0$$

Mean Squared Error is

$$E\left[(\hat{X}(h) - X_{n+h})^2\right]$$

95% prediction interval is

$$\hat{X}(h) \pm 1.96\sqrt{\hat{\mathrm{MSE}}}$$

Example: 10-step prediction

```
#--- When true mean is 0 ---
    X <- arima.sim(list(ar = c(.6, .3) ), 200 )

    Fit1 <- ar(X, aic=F, order.max=2, demean=FALSE )   #- Force to fit AR(2). don't de-mean.
    Fit1

    X.h  <- predict(Fit1, n.ahead=10)  #- 10-step prediction of AR(2)

    Xhat     <- X.h$pred
    Xhat.CIu <- Xhat+1.96*Y.h$se
    Xhat.CIl <- Xhat-1.96*Y.h$se

    ts.plot(cbind(Y, Yhat, Yhat.CIu, Yhat.CIl ), type="o", col=c("black","red","blue","blue"))
    abline(h=0)
    abline(h=mean(Y), col="blue")
```

```
#--- When true mean is not 0 ---
    Y    <- arima.sim(list(ar = c(.6, .3) ), 200 ) + 8

    Fit1 <- ar(Y, aic=F, order.max=2  )   #- Force to fit AR(2).  de-mean by default.
    Fit1

    Y.h  <- predict(Fit1, n.ahead=10)


    Yhat     <- Y.h$pred
    Yhat.CIu <- Yhat+1.96*Y.h$se
    Yhat.CIl <- Yhat-1.96*Y.h$se

    ts.plot(cbind(Y, Yhat, Yhat.CIu, Yhat.CIl ), type="o", col=c("black","red","blue","blue"))
    abline(h=0)
    abline(h=mean(Y), col="blue")
```

Rolling 1-step prediction (Fixed Window)

- Suppose you have $n = 100$ to begin.

  1. Use $\{X_1, \ldots, X_{100}\}$ to find fitting model. Say it was AR(2). Produce 1-step predicton $\{\hat{X}_{101}\}$
  2. Observe X_{101}
  3. Use $\{X_2, \ldots, X_{101}\}$ to re-estimate parameters of AR(2). Then produce 1-step predicton $\{\hat{X}_{102}\}$
  4. Observe X_{102}
  5. Use $\{X_3, \ldots, X_{102}\}$ to re-estimate parameters of AR(2). Then produce 1-step predicton $\{\hat{X}_{103}\}$
  6. Observe X_{103}

  $\vdots$

- Always using past 100 observation (Fixed Window) to predict the next obervation

- Fixed Start Rolling 1-step prediction will use $\{X_1, \ldots, X_{105}\}$ to predict X_{106}.

Example: Rolling 1-step prediction

```
#--- Rolling 1-step Prediction w/ fixed window of 100

    Y    <- arima.sim(list(ar = c(.6, .3)),  n=200, sd=.1  )    #- simulate the dataset

    plot(Y, type="o", col="blue")  #- Entire dataset

    X    <- Y[1:100]   #- First 100 obs
    lines(X, type="o")


    Fit1 <- ar(X)                         #- Find best model by AIC
```

```
#--- Suppose you are happy with AR(2).  Perform rolling 1-step prediction with widnow of 100 obs.
#--- AR(2) is fixed as rolled, but parameters are re-estimated each time.

    Yhat <- 0                       #- initialize
    Yhat.CIu <- 0
    Yhat.CIl <- 0
    for (i in 1:100) {

      window.bgn <- i
      window.end <- i+99

      Fit1 <- ar(Y[window.bgn:window.end], aic=FALSE, order.max=2 )     #- Force to fit AR(2) on last 100 obs.

      Y.h  <- predict(Fit1, n.ahead=1)

      Yhat[i]     <- Y.h$pred
      Yhat.CIu[i] <- Yhat[i]+1.96*Y.h$se
      Yhat.CIl[i] <- Yhat[i]-1.96*Y.h$se

    }
```

```
#--- Plot the prediction result with original data ---

    layout(matrix(c(1,1,1,2,2,3), 2, 3, byrow=TRUE))

    #- layout(1)  #- This turns off layout

    plot(Y, type="o", col="blue", main="Rolling 1-step prediction with window=100" )  #- Entire dataset
    lines(X, type="o")
    lines(101:200, Yhat, type="o", col="red")
    lines(101:200, Yhat.CIu, type="l", col="red", lty=2)
    lines(101:200, Yhat.CIl, type="l", col="red", lty=2)

    Pred.error = Y[101:200]-Yhat

    plot(101:200, Pred.error, type="o", main="Prediction Error (Blue-Red)")

    Pred.rMSE =  sqrt(  mean( (Pred.error)^2 ) )     #- prediction root Mean Squared Error
    abline(h=c(-1.96, 1.96), col="blue", lty=2)
    Pred.rMSE
    mean(Pred.error)


    acf(Pred.error)
```

5.3 Summary 5

[ToC]

---

1. Best 1-step linear predictor of AR(p) was simply to use the AR equation without e_t,

$$\hat{X}_{n+1} = \phi_1 X_n + \cdots \phi_p X_{n-p+1}$$

2. Looking h-step ahead, the "best prediction", value will dacay toward $E(e_t) = 0$ as h increases.

3. 1-step prediction's large-sample MSE is

$$\text{MSE} = \sigma^2$$

   where σ^2 is the variance of the error (innovation), e_t.

4. This leads to 1-step 95% prediction interval of

$$\hat{X}(h) \pm 1.96\sqrt{\text{MSE}}$$

5. One can perform 1-step rolling prediction to keep updating your forecast, as new observation arrives.

Additional Topic

- Linear Process

[ToC]

6.1 Linear Process

[ToC]

- To calculate ACVF of AR(p), it is easier to use the framework of Linear Process.

- Process is called linear process if you can write as

$$
\begin{aligned}
Y_t &= \sum_{i=-\infty}^{\infty} \psi_i e_{t-i} \\
&= \Psi(B)\, e_t \qquad\qquad \text{where } e_t \sim WN(0, \sigma^2)
\end{aligned}
$$

  for all t, and $\{\psi_i\}$ are sequence of constans that are absolutely summable.

- Note the sum goes from $-\infty$ to ∞, and this us sum of uncorrelated errors, e_t.

- The coefficients ψ_i must be absolutely summable, because it will imply that the sum is finite with probability one:

$$E|X_t| \leq \sum_{i=-\infty}^{\infty} |\psi_i| E|e_{t-j}| \leq \sigma \Big(\sum_{i=-\infty}^{\infty} |\psi_i| \Big)$$

- It also implies that ψ_i is square summable, and hence the partial sum converges to X_t in mean square.

$$X_t = \sum_{i=-\infty}^{\infty} \psi_i e_{t-i}$$

ACVF of Linear Process

can be calculated as

$$\begin{aligned}
\gamma(h) &= \mathrm{Cov}(X_{t+h},\, X_t) \\
&= \mathrm{Cov}\Big(\sum_{j=-\infty}^{\infty} \psi_j\, e_{t+h-j},\ \sum_{i=-\infty}^{\infty} \psi_i\, e_{t-i}\Big) \\
&= \text{ sum of cov of all possible pairs} \\
&= \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \mathrm{Cov}\Big(\psi_j\, e_{t+h-j},\ ,\psi_i\, e_{t-i}\Big) \\
&= \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \psi_j\, \psi_i\ \mathrm{Cov}\Big(e_{t+h-j},\, e_{t-i}\Big)
\end{aligned}$$

$$= \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \psi_j \; \psi_i \; \mathrm{Cov}\big(e_{t+h-j}, \, e_{t-i}\big)$$

$$= \sum_{i=-\infty}^{\infty} \; \psi_{i+h} \; \psi_i \; \mathrm{Cov}\big(e_{t-i}, \, e_{t-i}\big)$$

since e_i, e_j are independent for $i \neq j$. Therefore,

$$\gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_{i+h} \, \psi_i.$$

This equation gives us means to calculate ACVF of many time series.

6.2 Calculating ACVF of AR(p) using Linear Process

[ToC]

- Now we know how to calculate $\gamma(h)$ of any process that we can write as linear process. i.e. once we know ψ_i, we know ACVF.

- For AR(p), we can write it as causal process,

$$X_t = \sum_{i=0}^{\infty} \psi_i e_{t-i}$$

- Causal process **is** Linear Process. We just have all $\psi_{-1}, \psi_{-2}, \ldots$ equal to 0.

- So when we are given AR(p),

$$\begin{aligned}
\Phi(B)\, X_t &= e_t \\
X_t &= \frac{1}{\Phi(B)}\, e_t = \sum_{i=0}^{\infty} \psi_i\, e_{t-i} = \Psi(B)\, e_t
\end{aligned}$$

we need to figure out ψ_i, in terms of ϕ_i.

- We have equation,

$$\begin{aligned}
\frac{1}{\Phi(B)} &= \Psi(B) \\
\frac{1}{\left(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p\right)} &= \psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \cdots
\end{aligned}$$

We can do this by comparing the like terms.

$$1 = \left(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p\right)\left(\psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \cdots\right)$$

- That means, comparing coeficccients of each term of z,

$$1 = \left(1-\phi_1 z-\phi_2 z^2-\cdots-\phi_p z^p\right)\left(\psi_0+\psi_1 z+\psi_2 z^2+\psi_3 z^3+\cdots\right)$$

$$\begin{aligned}
1 &= \psi_0 && \text{coefficient without } z \\
0 &= \psi_1-\psi_0\phi_1 && \text{coefficient for } z \\
0 &= \psi_2-\psi_1\phi_1-\psi_0\phi_2 && \text{coefficient for } z^2 \\
0 &= \psi_3-\psi_2\phi_1-\psi_1\phi_2-\psi_0\phi_3 && \text{coefficient for } z^3 \\
&\vdots && \vdots
\end{aligned}$$

- We turn this around and write

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \psi_0\phi_1 \\
\psi_2 &= \psi_1\phi_1+\psi_0\phi_2 \\
\psi_3 &= \psi_2\phi_1+\psi_1\phi_2+\psi_0\phi_3 \\
&\vdots
\end{aligned}$$

Recursive formula for ACVF of AR(p)

- Now we know that AR(p) can be written in causal form (which is Linear Process) with coefficients,

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \psi_0\phi_1 \\
\psi_2 &= \psi_1\phi_1 + \psi_0\phi_2 \\
\psi_3 &= \psi_2\phi_1 + \psi_1\phi_2 + \psi_0\phi_3 \\
&\vdots
\end{aligned}$$

- In short,

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_i &= \sum_{k=1}^{p} \phi_k \psi_{i-k} \qquad\qquad i > 0
\end{aligned}$$

- Using this formula, we can numerically calculate the ACVF of AR(p).

Example: AR(2)

Let Y_t be the AR(2) process

$$Y_t = .7Y_{t-1} - .1Y_{t-2} + Z_t$$

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

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= .7 \\
\psi_2 &= .7^2 - .1 \\
&\vdots \\
\psi_i &= .7\psi_{j-1} - .1\psi_{j-2} \quad i \geq 3 \\
&\vdots
\end{aligned}$$

Then after getting ψ_i, calculate the ACVF by the formula

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}.$$

ACVF of AR(2)

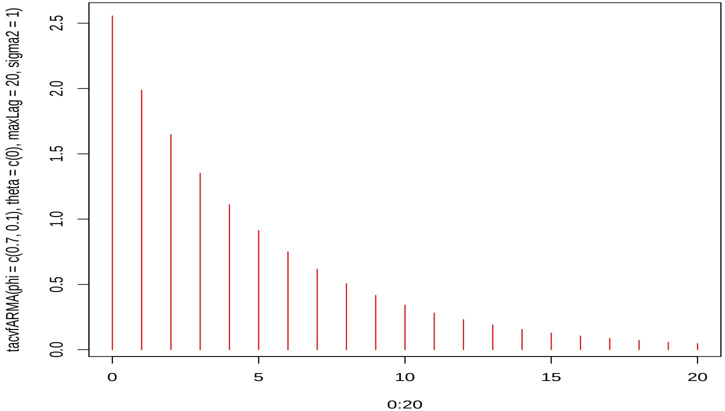
$\Phi(z) = 1 - .7z - .1z^2$

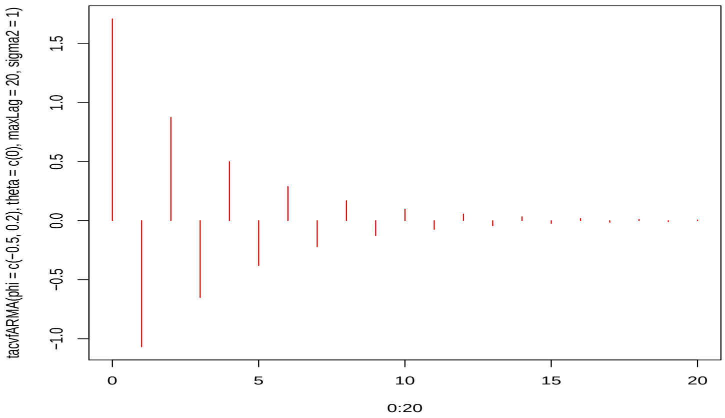
$\Phi(z) = 1 + .5z - .2z^2$

```
#- install.packages("ltsa")

library(ltsa)

plot(0:20, tacvfARMA(phi = c(.7, .1), maxLag=20, sigma2=1), type="h")
plot(0:20, tacvfARMA(phi = c(-.5, .2), maxLag=20, sigma2=1), type="h")
```


tacvfARMA(phi = c(0.7, 0.1), theta = c(0), maxLag = 20, sigma2 = 1)
0.0
0.5
1.0
1.5
2.0
2.5
0
5
10
15
20
0:20


tacvfARMA(phi = c(−0.5, 0.2), theta = c(0), maxLag = 20, sigma2 = 1)
-1.0
-0.5
0.0
0.5
1.0
1.5
0
5
10
15
20
0:20

6.3 Large sample property of sample ACF

[ToC]

- Bartlett's Formula

$$\hat{\boldsymbol{\rho}} \sim \mathcal{N}\Big(\boldsymbol{\rho}, \frac{1}{n}\boldsymbol{W}\Big)$$

$$\begin{aligned}
\boldsymbol{W} &= \Big\{w_{ij}\Big\} \\
w_{ij} &= \sum_{k=1}^{\infty}\Big[\rho(k+i)+\rho(k-i)-2\rho(i)\rho(k)\Big] \\
&\qquad\qquad \cdot\Big[\rho(k+j)+\rho(k-j)-2\rho(j)\rho(k)\Big]
\end{aligned}$$

under iid, $\boldsymbol{W}$ becomes identity matrix.

Example: MA(1)

$$w_{ii} = \begin{cases} 1 - 3\rho^2(1) + 2\rho^4(1) & \text{if } i = 1 \\ 1 + 2\rho^2(1) & \text{if } i > 1 \end{cases}$$

Example: AR(1)

Recall that $\phi(h) = \phi^{|h|}$. We have

$$w_{ii} = (1 - \phi^{2i})(1 + \phi^2)(1 - \phi^2)^{-1} - 2i\phi^{2i}$$