Tutorial 2

This tutorial deals with ARMA processes, their autocovariance and autocorrelation functions, their partial autocovariance. If $\rho(h) = 0$ for $|h| > q$ (autocorrelation), then an MA(q) process may be a good fit; if $\alpha(h) = 0$ for all $|h| > p$ (partial autocorrelation) then an AR(p) process may be a good fit.

Note An ARIMA(p,d,q) process is an *integrated* ARMA process. $\{X_t\}$ is said to be ARIMA(p,d,q) if $\{\nabla^j X_t\}$ is *not* stationary for $j < d$, but $\{\nabla^d X_t\}$ is an ARMA(p,q) process. X is an ARMA(p,q) process if it can be written in the form:

$$X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \qquad \{\epsilon\} \sim WN(0, \sigma^2).$$

Introductory

A *correlogram* is a plot of the autocrrelation function. Recall that for the AR(1) process, the autocorrelation is $\rho(k) = \phi^k$. Firstly, plot the correlograms for $\phi = 0.7$ and $\phi = -0.7$.

```
library(ggplot2)
require(gridExtra)
rho <- function(k, alpha) alpha^k
plot1 = qplot(0:10, rho(0:10, 0.7),geom="path")
plot2 = qplot(0:10, rho(0:10, -0.7),geom="path")
grid.arrange(plot1, plot2, nrow=2)
```

Try experimenting for other values of ϕ.

Now simulate an AR(1) process:

```
set.seed(1)
x <- w <- rnorm(100)
for (t in 2:100) x[t] <- 0.7 * x[t - 1] + w[t]
plot(x, type = "l")
acf(x)
pacf(x)
```

the `pacf` is the Partial Autocorrelation Function.

There are several ways to fit an AR model to data; one of the most effective is `auto.arima`. We'll use this on `x`. The order of the estimated model is extracted using `$order` and the variance of the parameter estimate using `$asy.var`:

```
x.ar = auto.arima(x, max.q=0, stationary=TRUE, seasonal=FALSE)
> x.ar
Series: x
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1    mean
      0.6010  0.3544
s.e.  0.0808  0.2196

sigma^2 estimated as 0.8042:  log likelihood=-130.21
AIC=266.42   AICc=266.67   BIC=274.24
```

Now try fitting an ARIMA model to the New-Zealand exchange rate (to pounds sterling) data:

```
www <- "http://www.mimuw.edu.pl/~noble/courses/TimeSeries/data/pounds_nz.dat"
Z <- read.table(www, header = T)
Z.auto <- auto.arima(Z)
Z.auto
Series: Z
ARIMA(0,1,1)

Coefficients:
         ma1
      0.4771
s.e.  0.1973

sigma^2 estimated as 0.0163:  log likelihood=24.67
AIC=-45.35   AICc=-45.01   BIC=-42.07
```

An integrated MA(1) model fits.

Now, note that the series is not mean zero:

```
> mean(Z$xrate)
[1] 2.823251
> sd(Z$xrate)
[1] 0.3829946
```

In fact, the mean is very substantial compared with the standard deviation. We can allow a non-zero mean by:

```
Z2.auto=auto.arima(Z,allowmean=TRUE)
Z2.auto
Series: Z
ARIMA(0,1,1)

Coefficients:
         ma1
      0.4771
s.e.  0.1973

sigma^2 estimated as 0.0163:  log likelihood=24.67
AIC=-45.35   AICc=-45.01   BIC=-42.07
```

which gives the same answer; `auto.arima` chooses the zero mean model.
We can insist on stationarity. Let us fit an AR model:

```
Z3.ar = auto.arima(Z,max.q=0,allowmean=TRUE,stationary=TRUE)
Z3.ar
Series: Z
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2    mean
      1.3091  -0.3857  2.9457
s.e.  0.1475   0.1498  0.2334

sigma^2 estimated as 0.01672:  log likelihood=24.72
AIC=-41.44   AICc=-40.27   BIC=-34.79
```

By default, the mean is subtracted before the parameters are estimated so the estimated model is:

$$\widehat{x}_t = 2.95 + 1.3(x_{t-1} - 2.95) - 0.39(x_{t-2} - 2.95) + \epsilon_t.$$

The two models perform similarly; the σ^2 estimate is similar. Preference is therefore given to the ARIMA(0,1,1) model, since it has fewer parameters.

Global Temperature Series The data set `global.dat` gives the global temperatures and indicate an increasing trend after 1970. The aim of the following exercise is to decide whether or not this is a transient stochastic phenomenon.

The data points are the monthly temperatures. We aggregate over the whole year. A plot of the data clearly indicates a pronounced *trend*, yet we may apply the algorithm for fitting an AR process,

```
glob<-"https://www.mimuw.edu.pl/~noble/courses/TimeSeries/data/global
.dat"
Global<-scan(glob)

Read 1800 items

Global.ts=ts(Global,fr=12)
Global.ar <- ar(aggregate(Global.ts, FUN = mean), method = "mle")

Warning message:
In arima0(x, order = c(i, 0L, 0L), include.mean = demean) :
  possible convergence problem: optim gave code = 1

mean(aggregate(Global.ts,FUN=mean))
[1] -0.1382628

Global.ar$order
[1] 4

Global.ar$ar
[1] 0.58762026 0.01260253 0.11116731 0.26763656

acf(Global.ar$res[-(1:Global.ar$order)], lag = 50)
```

The correlogram is 'white noise', so in principle, an AR(4) model *could* be used explain the data, if it were not very clear that there is an increasing trend.

Exercises

1. (a) Simulate a time series, length 1000, for the model:

$$x_t = \frac{5}{6}x_{t-1} - \frac{1}{6}x_{t-2} + Z_t$$

   (b) Plot the acf and partial acf for the simulated data. Comment on the plots.

   (c) Fit an AR model to the simulated data giving the parameter estimates and the order of the fitted AR process.

   (d) Construct 95% confidence intervals for the parameter estimates of the fitted model. Do the model parameters fall within the confidence intervals? Explain your results.

   (e) Is the model stationary or non-stationary? Justify your answer.

   (f) Plot the acf of the residuals of the fitted model. Comment on the plot.

2. (a) Refit the AR(4) model to the global temperature data and use the fitted model to create a series of predicted values from $t = 2$ to the last value in the series. Create a residual series from the difference between the predicted values and the observed value and verify that the series is identical (within computational accuracy) to the series extracted from the fitted model in R.
   (b) Plot the acf and pacf for the mean annual temperature series. Comment on the plots.
   (c) Use the `predict` function in R to forecast 100 years of future values for the annual global temperature series using the fitted AR(4) model.
   (d) Create a time plot of the mean annual temperature series and add the 100-year forecasts to the plot. Use a different symbol or colour for the forecasts.
   (e) Add a line representing the overall mean global temperature. Comment on the final plot and any potential inadequacies in the fitted model.

3. **Deterministic Seasonal Behaviour**

The data set `m-deciles08.txt` contains the monthly simple returns of the CRSP Decile 1 Index from January 1970 to December 2008, for 468 observations. First, make time series plot. Is there any apparent seasonality? What is the sample ACF? What are the significant lags?

A seasonal ARMA model for a process with lags at 1, and $12k$ for $k \geq 1$, would be:

$$(1 - \phi_1 B)(1 - \phi_{12} B^{12}) X_t = (1 - \theta_{12} B^{12}) Z_t \qquad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$$

What are the parameter estimates ϕ_1, ϕ_{12} and θ_{12} for such a model?

The seasonal behaviour might be deterministic. Try to create a dummy variable

$$J = \begin{cases} 1 & \text{January} \\ 0 & \text{otherwse} \end{cases}$$

and make a regression

$$X_t = \beta_0 + \beta_1 J_t + \epsilon_t.$$

What is the fitted model? Does it fit the data well? What do you conclude about the 'January effect?'

The following may be useful.

```
#load the data set, call the table m.deciles08
da = m.deciles08
d1 = da[,2]
jan=rep(c(1,rep(0,11)),39) #create January dummy
m1 = lm(d1~jan)
summary(m1)
```