

APTS Statistical Modelling: Practical 1

D. C. Woods and A. M. Overstall

March 24, 2015

The code below generates a time series of length n , and then fits autoregressive models of order up to `order.max` (which is set to 19 below). The AIC is plotted as a function of the order, and the optimal order is tabulated.¹

```
plot.aic <- function(fit, new=T, sd=0.1)
{ # code to plot AIC against order of AR model fitted
if (new) plot((1:length(fit$aic))-1,fit$aic,type="l",xlab="Order",ylab="AIC") else
    lines((1:length(fit$aic))-1,fit$aic,type="l")
    points(rnorm(1,fit$order,sd=sd),rnorm(1,sd=sd),pch=16,col="red")
}

# generates data from an autoregressive process, by default of order 1
sim.y <- function(n, model=list(ar=c(0.9))) arima.sim(model=model, n)

n <- 25 # length of time series
R <- 1000 # number of replicates

# first dataset to get things started
y <- sim.y(n,list(ar=c(0.5,0.1)))
fit <- ar(y,order.max=19)
plot.aic(fit)

# we will store the orders chosen using AIC, BIC, and AICC
AIC.order <- NULL
BIC.order <- NULL
AICC.order <- NULL

# Now make R replicates, plot the corresponding AIC curves
for (i in 1:R )
{
fit <- ar( sim.y(n, model=list(ar=c(0.5,0.1))), order.max=19 )
plot.aic(fit, new=F)
AIC.order <- c(AIC.order, fit$order)
# The next two lines should be uncommented and modified to give the
# optimal orders when BIC and AICC are used for order selection
# BIC.order <- c(BIC.order, NA)
# AICC.order <- c(AICC.order, NA)
}

# tabulate the order of the chosen model
table(AIC.order)
```

¹The code is available from the APTS website.

- (a) Try seeing how AIC performs as a basis for model selection for $n = 25, 50, 100, 500$.
- (b) Vary the simulation model, using, for example, `model=list(ma=0.9)` in the `arma.sim` function, to see how well AIC works when the data are not generated by an autoregressive model.
- (c) Modify the code above to compute the values of BIC and AIC_c , where

$$AIC = 2(p + 1 - \hat{\ell}), \quad BIC = p \log n - 2\hat{\ell}, \quad AIC_c = n \frac{n + p}{n - p - 2} - 2\hat{\ell},$$

and p is the order of the fitted model. Assess how well these criteria perform as bases for model selection, for $n = 25, 50, 100, 500$.

Hint: write BIC and AIC_c as functions of AIC. You may find it useful to use `?ar` to access the help file for the `ar` function for details of the returned elements (or use `ls(fit)` to list them).