Statistical Modelling: Practical 1

A. C. Davison and J. J. Forster

April 8, 2008

The code below generates a time series of length n, and then fits autoregressive models of order up to `order.max`, which is 20 below. The function AIC is plotted as a function of the order, and the optimal order is tabulated.[1]

```
plot.aic <- function(fit, new=T)
{  # 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(fit$order,0,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 <- 20  # length of time series
R <- 1000  # number of replicates

# first dataset to get things started
y <- sim.y(n)
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))) )
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)
```

[1] The code is available from the APTS website.

(a) Try seeing how AIC performs as a basis for model selection for $n = 20, 50, 100, 500$.

(b) Vary the simulation model, using the `ma` argument in order for the `arima.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

$$\text{AIC} = 2(p + 1 - \widehat{\ell}), \quad \text{BIC} = p \log n - 2\widehat{\ell}, \quad \text{AIC}_\text{c} = n\frac{n+p}{n-p-2} - 2\widehat{\ell},$$

where p is the order of the fitted model, and assess how well they perform as bases for model selection, for $n = 20, 50, 100, 500$.

(d) If you have time, include AIC_u in (c), with $\widehat{\sigma}^2$ replaced by the unbiased estimate of variance.

(e) If you are not yet bored, reproduce the figures given in the lectures (slides 37–39), for this setting. Comment.