APTS Statistical Modelling: Practical 1

A. M. Overstall

March 2017

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$order, sd = sd), rnorm(1, sd = sd), pch = 16,
+         col = "red")
+ }
>
> # generates data from the model given by model
> sim.y <- function(n, model) {
+     arima.sim(model = model, n)
+ }
>
> n <- 25  # length of time series
> R <- 1000  # number of replicates
>
> # first dataset to get things started it is an AR process of
> # order 2.
> y <- sim.y(n, list(ar = c(0.5, 0.1)))
> fit <- ar(y, order.max = 19)
> plot.aic(fit)
>
> # we will store the orders chosen using AIC, BIC, and AICC
> AIC.order <- NULL
> BIC.order <- NULL
> AICC.order <- NULL
>
> # Now make R replicates, plot the corresponding AIC curves
> for (i in 1:R) {
+     y <- sim.y(n, list(ar = c(0.5, 0.1)))
+     fit <- ar(y, order.max = 19)
+     plot.aic(fit, new = F)
+     AIC.order <- c(AIC.order, fit$order)
+     # The next two lines should be uncommented and modified to
+     # give the optimal orders when BIC and AICC are used for
```

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

```
+     # order selection BIC.order <- c(BIC.order, NA) AICC.order <-
+     # c(AICC.order, NA)
+ }
>
> # tabulate the order of the chosen model
> table(AIC.order)
```

(a) Try seeing how AIC performs as a basis for model selection for $n = 25, 50, 100, 1000$.

(b) Vary the simulation model, using, for example, `model=list(ma=0.9)` in 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

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

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).