Stat 510, Lab 6

October 15, 2010

1 In Class

1. In this lab, we will learn how to do maximum likelihood estimation for time series using R. We will start with a data set of sulfur dioxide levels in LA at weekly time intervals from 1970-1979. Download and input the data with the following commands

   ```
   x=scan("so2.dat")
   ```

2. Check out plots of the data and the differenced data–plot, acf, and pacf.

3. We can fit models using numerical maximum likelihood estimation. We will be using a function in R to fit our data–to fit an $ARMA(2,1,1)$, we use `fit1=arima(x,c(2,1,1))` to obtain a fit of the data. We type `fit1` to obtain the following:

   Let us look at the output.

   ```
   Call:
   arima(x = x, order = (c(2, 1, 1)))

   Coefficients:
            ar1     ar2      ma1
         0.0111  0.1100  -0.8777
   s.e.  0.0684  0.0634   0.0488

   sigma^2 estimated as 0.7785:  log likelihood = -656.57,  aic = 1321.13
   ```

   Notice that we have estimate for the mean and for ϕ_1, ϕ_2, and θ_1 along with standard errors. We also have an estimate for the variance of the

white noise, σ^2. We also have the AIC which allows us to compare different models; test this out with a few other models.

We can obtain diagnostics of the fit with the command `tsdiag(fit1)`. This give us the standardized residuals and the ACF of the residuals. These should appear as white noise.

4. Now, let's load and work with the summer temperature data. Use the following command to input the data

```
summer=scan("summer.dat", skip=1)
```

This file contains the mean summer temperature in Munich from 1781-1988. Use the `arima()` function to fit ARIMA models to this data.

2 Homework

This week we will be analyzing a number of data sets. We are going to build ARIMA models using the steps outlined in class. It is also a good idea to read section 3.8 from our textbook. Make sure to outline the steps used in analyzing the data. If there are two (or more) competing models, make sure you discuss each of these. Make a decision about which model you think is best and support this decision with plots and and other information. I suggest using the `arima()` function to do these fits. Here are short summaries about the data sets to analyze.

1. The data in `cow.dat` are the daily morning temperature readings for a cow.

2. The data in `sheep.dat` are the sheep population (in millions) for England and Wales from 1867-1939.

3. The data in `bicoal.dat` are the annual bituminous coal production levels in the US from 1930-1968 in millions of net tons per year.