STAT 443 – Winter 2015 – Assignment 1

due Thursday January 29 at the beginning of class

You may work in pairs if you choose; both names and ID numbers should appear on it, and both will receive the same mark. (No extra credit will be given for working alone.)

For any parts involving R, you should hand in the R code and output, as well as your interpretations of the output. You will NOT receive marks for uncommented R code or output.

1. Consider a second-order moving average (MA(2)) process defined by $X_t = Z_t + \theta Z_{t-1} + \theta^2 Z_{t-2}$, where $\{Z_t\} \sim WN(0, \sigma^2)$. Assume that θ and σ are finite.

    (a) Show that this process is stationary.

    (b) What are the maximum and the minimum values of the autocorrelation function of X_t at lag 2, $\rho(2)$? At what values of θ are the maximum and minimum attained?

    (c) Use R to generate a sample of 100 realizations of an MA(2) process with $\{Z_t\} \sim$ *iid* $N(0, 1)$ and each of the following values of θ: -1, -0.2, 0.5, 0.9. Plot the sample autocorrelation function (SACF) of each one, and comment on the plots.

2. Consider the first-order auto-regressive (AR(1)) process defined by $X_t = \phi X_{t-1} + Z_t$, where $\{X_t\}$ is stationary, $\{Z_t\} \sim WN(0, \sigma^2)$, and Z_r and X_s are independent for all $r > s$. Prove directly that if $h < 0$, $\gamma(h) = \phi^{-h}\sigma^2/(1 - \phi^2)$.
Hint: use a very similar trick to the one we used in class for $h > 0$. Be sure to explain all your steps clearly.

3. Let $\{X_t\}$, $\{Y_t\}$, and $\{Z_t\}$ be three independent stationary time series with means μ_X, μ_Y, and μ_Z and autocovariance functions (ACVFs) γ_X, γ_Y, and γ_Z, respectively. The time series $\{Z_t\}$ takes only two values: 0 and 1.
Let $U_t = \begin{cases} X_t & \text{if } Z_t = 0 \\ Y_t & \text{if } Z_t = 1 \end{cases}$ or, equivalently, $U_t = (1 - Z_t)X_t + Z_t Y_t$ for all t.
Show that the time series $\{U_t\}$ is stationary.

4. Use R to generate samples of 100 realizations (for $t = 1, ..., 100$) of each of the following NON-stationary time series. For each one, plot the data and the SACF and comment on the plots, including how you can tell stationarity is violated and the impact of the parameters.

    (a) $X_t = 22 - 15t + 0.3t^2 + Z_t$, where $\{Z_t\} \sim iid\ N(0, 50^2)$

    (b) $X_t = X_{t-1} + Z_t$, where $\{Z_t\} \sim iid\ N(0, 1)$ and start with $X_0 = 0$

    (c) $X_t = 22 - 1.5t + 0.01t^2 + 5sin(2t) + Z_t$, where $\{Z_t\} \sim iid\ N(0, 2^2)$

    (d) $X_t = sin(t)$

    (e) $X_t = sin(t) + Z_t$, where $\{Z_t\} \sim iid\ N(0, 0.5^2)$

    (f) $X_t = sin(t) + Z_t$, where $\{Z_t\} \sim iid\ N(0, 1)$

    (g) $X_t = sin(t) + Z_t$, where $\{Z_t\} \sim iid\ N(0, 2^2)$

5. The file lightning.txt includes the exact number of lightning strikes per month observed in a cirlce around an Ontario power plant (located at approximately 44.3 degrees North, 81.6 degrees West) from September 2001 to September 2011. Consider the first 9.5 years (114 months) as the training set and the last 7 months as the testing set. (I would like to thank Dr. Bill Chisholm, retired from Ontario Hydro, for the data.)

   To read this data into R and turn it into a time series format, use the following commands:

```
lightning <- read.table('lightning.txt')
light.all <- ts(lightning, start = c(2001,9), frequency = 12)
light.train <- ts(lightning[1:114,1], start = c(2001,9), frequency = 12)
light.test <- ts(lightning[115:121,1], start = c(2011,3), frequency = 12)
```

   (a) Fit a seasonal (no trend) regression model ($L_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + ... + \beta_{11} X_{t,11} + \epsilon_t$, where the $X_{t,j}$'s are indicator variables for the jth month) to the traning set data and comment on the fit of the model. You should fit the model (provide your commented R code along with the R output), and provide full residuals diagnostics (comment on the 4 plots discussed in class as well as the results of the hypothesis tests discussed in class).

   (b) Using the model, predict the number of lightning strikes in each of the last 7 months (i.e. March 2011 to Sept 2011) and provide 95% prediction intervals. Compare the actual data for those months to your intervals. Comment on the performance of your forecasting model.

   (c) Because of the large and fluctuating variance, you decide to perform a variance-stabilizing transformation and re-fit the model. Unfortunately there are 0's in the data so you cannot use the log transformation. Instead, use $\log(L_t + 1)$. (This has the advantage that 0 is transformed to 0.) Re-fit the model, diagnostics, and predictions as in parts (a) and (b) (remember to un-transform to get the prediction intervals).

   (d) Compare the fit and performance of the two models. Which, if any, satisfies the fundamental assumptions of a regression model?