# Stat 5170: Assignment 6

due Mar 25 2016, 5pm

## 1 R Tutorial

1. Reading R code from a file. I created a function called `armasim` to simulate ARMA processes. Download and save the file "armasim.R" in your working directory. Use the `source` function to read R code from a file, e.g.: `source("armasim.R")`. You have to specify separate vectors for the parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, the variance of the white noise terms, and the length of the simulation $T$.

   (a) Obtain the ACF and PACF plots for the following process:

   $$x_t = 0.5x_{t-1} + w_t - 0.5w_{t-1}.$$

   Use the following code:

   ```
   source("armasim.R")
   x1<-armasim(c(0.5),c(-0.5),1,10000)
   ```

   You may use the `armasim()` function to generate AR(p) and MA(q) processes. Just be sure to enter `0` for the vector of $\boldsymbol{\phi}$ or $\boldsymbol{\theta}$ in the function.

   (b) Scale of axes. The `acf()` function in R produces plots of the sample ACFs from **lag 0**. On the other hand, the `pacf()` function in R produces plots of the sample PACFs from **lag 1**. As a result, the scales in the plots are often very different and may be misleading. To have the ACF plot start from lag 1, and have the same vertical scale for both the ACF and PACF plots, we could use, for example,

   ```
   par(mfrow=c(1,2))
   acf(x1, 20, xlim=c(1,20), ylim=c(-0.03,0.03), main="ACF")
   pacf(x1, 20, ylim=c(-0.03,0.03), main="PACF")
   ```

   Compare with

   ```
   par(mfrow=c(1,2))
   acf(x1, main="ACF")
   pacf(x1, main="PACF")
   ```

## 2 Assignment

1. Use the `armasim()` function provided to simulate ARMA(p,q) processes.

   Explore the behavior of the ACF and PACF. Simulate a single path of 10,000 observations for each of the following models:

   (a) $x_t = 0.6x_{t-1} - 0.8x_{t-2} + w_t$

   (b) $x_t = w_t + 0.8w_{t-1} + 1.1w_{t-2}$

   (c) $x_t = 0.8x_{t-1} + w_t + 0.8w_{t-1}$

   Plot the ACF and PACF for each process. Discuss what you see in the plots and if it is what you would expect.

2. Load the Berkeley and Santa Barbara temperature data using the following commands

```
berk<-scan("berkeley.dat", what=list(double(0),double(0),double(0)))
time<-berk[[1]]
berkeley<-berk[[2]]
stbarb<-berk[[3]]
```

   Create ACF and PACF plots for `berkeley` and `stbarb`. Repeat this for the differenced data. Do you have an opinion on possible AR and/or MA models based only on these plots?

3. (No R required) Consider an AR(1) process of the form $x_t = \phi x_{t-1} + w_t$. Let $\sigma_w^2$ denote the variance of $w_t$. You may assume this is a zero mean process, that $\gamma_x(h) = \sigma_w^2 \frac{\phi^{|h|}}{1-\phi^2}$, and that $\rho_x(h) = \phi^{|h|}$.

   (a) Based on the best linear predictor, write the prediction equations for an AR(1) process.

   (b) Show that the general form of the $m$-step-ahead forecast is $x_{t+m}^t = \phi^m x_t$.

   (c) Show that the mean squared error is $\mathrm{E}\left[(x_{t+m} - x_{t+m}^t)^2\right] = \sigma_w^2 \frac{1-\phi^{2m}}{1-\phi^2}$.

4. (No R required for part 4b to 4f) Suppose that annual sales (in millions of dollars) of the Acme Corporation follow an AR(2) model: $x_t = 5 + 1.1x_{t-1} - 0.5x_{t-2} + w_t$, with $\sigma_w^2 = 2$.

   (a) Verify that this is a causal AR(2) model (you may use R to find the roots).

   (b) If sales for 2005, 2006, and 2007 were \$9 million, \$11 million, and \$10 million, respectively, using one-step ahead prediction, forecast sales for 2008 and 2009.

   (c) Calculate 95% prediction intervals for your forecasts in part 4b.

   (d) Briefly explain, in words, why the prediction interval for the forecast in 2009 is wider than the prediction interval for the forecast in 2008.

(e) If the sales in 2008 turned out to be \$12 million, would this value surprise you?

(f) Since we now know the sales for 2008 is \$12 million, update your forecast for the year 2009.