1. Suppose that $\{e_t\}$ is a zero mean white noise process with $\text{var}(e_t) = \sigma_e^2$, and consider the AR(1) model with $\phi = 0.8$, that is,

$$Y_t = 0.8Y_{t-1} + e_t.$$

Suppose that we have observed a series of length $n = 200$ from this AR(1) process.
(a) Write out the approximate (large-sample) sampling distributions of r_1, r_2, r_5, and r_{10}.
(b) Use Monte Carlo simulation, like that in Example 6.2 (notes), to simulate the sampling distributions of r_1, r_2, r_5, and r_{10} when $n = 200$. Assume that $e_t \sim \text{iid } \mathcal{N}(0, 1)$. Do these sampling distributions agree with what the large-sample theory says should happen?

2. In Chapter 6, we described testing hypothesis testing procedures to select MA and AR models using sample autocorrelations and sample partial autocorrelations, respectively. Use these testing procedures in each of the following situations.
(a) From a time series of $n = 150$ observations, we compute the following sample autocorrelation coefficients: $r_1 = -0.37$, $r_2 = 0.28$, $r_3 = 0.31$, $r_4 = -0.13$, $r_5 = 0.04$, and $r_6 = -0.15$. The remaining sample autocorrelations appear to be negligible. Find the MA(q) process most consistent with this information, that is, determine the value of q.
(b) From a time series of $n = 150$ observations, we compute the following sample partial autocorrelation coefficients: $\widehat{\phi}_{11} = 0.24$, $\widehat{\phi}_{22} = -0.81$, $\widehat{\phi}_{33} = 0.31$, $\widehat{\phi}_{44} = -0.09$, $\widehat{\phi}_{55} = 0.04$, and $\widehat{\phi}_{66} = -0.02$. The remaining sample partial autocorrelations appear to be negligible. Find the AR(p) process most consistent with this information, that is, determine the value of p.

3. Generate three time series data sets, each of length $n = 200$, including (i) an AR(1) with $\phi = -0.6$, (ii) an MA(1) with $\theta = 0.8$, and (iii) an ARMA(1,1) with $\phi = -0.6$ and $\theta = 0.8$. For each one,

(a) plot the observed time series.

(b) plot the sample ACF, the sample PACF, and the sample EACF.

(c) use the `armasubsets` function in R to identify the best model in terms of the BIC.

Do the plots in part (b) agree with what you know to be true? Remember, you know the correct models! That is, you are assessing here whether the sample identification functions agree with the truth. Does the BIC identify the correct model as the "best" model in each case? If not, where is the correct model ranked, if at all?

4. I have put two new data sets on the course web site:

- `ibm`: daily closing IBM stock prices (dates not given)

- `internet`: number of users logged on to an Internet server each minute.

In addition, consider the data set

- `robot`: final horizontal position of an industrial robot put through a series of planned exercises

which is in the `TSA` package.

Using the methods from Chapter 6, identify a small set of candidate ARIMA(p, d, q) models for each data set. There may be a single model that emerges as a "clear favorite." For guidance, use the summary described in Section 6.7 (notes) and follow it exactly. For each data set, write up detailed notes that describe how you decided on the model(s) you did. Your summary should convince me that your model(s) is (are) worthy of further consideration.

Note: Problem 4 is important because your class project will involve you finding data of your own, specifying a model (or models), fitting the model(s), and diagnosing model fit. This problem will help you in the model specification phase of your project.

Important: Remember your selected candidate model(s) in Problem 4 for each data set. You will consider these models further in Homework 5.