Economics 403A: Project 1
Fall 2017, UCLA
Instructor: Dr. Rojas

Due Date: October 19, 2017

The document that you will submit, consists of a written report which includes answers to the questions below (including plots), and respective R source code.

1. Fitting Distributions (30%)
   In this portion, you will be looking at different macroeconomic variables and determining whether or not the underlying distribution of these variables changes depending on whether or not the economy is in a state of recession. The variables you will be using will be the following:

   - S&P 500 index
   - Yield Spread (For simplicity, compute the spread as the difference between the ten year treasury bill rate and the three month treasury bill rate.)
   - Three month treasury rate
   - Japanese Central Bank's interest rates

   Use monthly frequencies for the data. At a minimum, your data should span at least 50 years.

   (a) For each of the four variables, plot the histogram of the data and overlay the respective density curve.

   (b) Fit a distribution to each histogram in part (a).

   (c) Now, subset the data to include only data points generated during recession periods. Repeat parts (a) and (b) with this subset of data.
   Note: for recession periods, use the National Bureau of Economic Research's classification for when the United States is in a recession. For data corresponding to Japan, classify recessions in accordance with the Organization for Economic Cooperation and Development (OECD).

   (d) Repeat part (c) with data subsetted to include only data points generated during non-recession periods.

   (e) Based on your findings in parts (a) - (d), are there any noticeable differences between the estimated distributions based on how you subset the data? Explain.

   (f) Confirm your answer from part (e) using the Kolmogorov-Smirnov Test.

2. Characterizing Financial Data (30%)

   (a) Choose four stocks (any that you like) and plot their historical prices (use daily frequencies).

   (b) For each stock compute the relevant summary statistics (mean, standard deviation, median, Q_1, Q_3, and correlation with the S&P500), and show your results in a table.

   (c) Plot a single figure showing the boxplots of all 4 stocks and the S&P500.

   (d) Convert the prices to daily returns, and plot the respective histograms and density curves.

   (e) Compute the excess return of each stock with respect to the S&P500. Note: The excess return can be computed by simply taking the difference between the stock returns (R_{stock}) and the S&P500 returns ($R_{S\&P500}$), i.e., $R_{Excess} = R_{stock} - R_{S\&P500}$. Fit a distribution to the excess returns, and compute the probability that the respective returns exceed the S&P500 returns by 3%.

   (f) Based on your findings, would you recommend investing in these stocks?

3. Sequential Bayesian Learning (30%)

   You will need to implement this in R using loops. Assume we are given the following facts:

   - 1% of women aged 40 have breast cancer.
   - A mammography test has a 99% success rate, and a 10% false alarm rate (i.e., 10% of the time, the test will return positive for having cancer when the patient does not actually have cancer).

   (a) Given the above, a women aged 40 receives a positive mammography test. What is the probability that she actually has cancer? Let $C+$ = Cancer present and $T+$ = Positive mammography test, you need to find$P(C+|T+)$.

   (b) Assume the same women from (a), wants to get another opinion, and then another, and so on, but that every time she gets tested, the results are the same as the first one. After how many trials, will $P(C+|T+) > 95\%$? Show a plot of $P(C+|T+)$ vs Trial Number, and comment on your finding.

   (c) If on Trial 3, the test results show negative for breast cancer, but all other tests show positive, how would this affect $P(C+|T+)$ found in (b)?

4. R Source Code (10%)
   Your code needs to include proper comments to help e.g., a non-R expert understand and run your code. If you do not submit your code, you will not receive credit for the project.