Homework 6

46-923, Fall 2017

Due Thursday, December 14 at 3:00 PM

You should complete this homework solely within R Markdown. You do not need to show any derivations that were required to complete Question 1, just implement the analysis in R.

Question 1

Assume that $X_1, X_2, \ldots, X_n$ are iid from the Poisson(λ) distribution.

Assume you want to test $H_0 : \lambda = 4.3$ versus $H_1 : \lambda < 4.3$ using the Wald test. Calculate the p-value when the data are those found via

```
x = read.table("http://www.stat.cmu.edu/~cschafer/MSCF/PoisHypTest.txt")

## Fetch data mean (lambda_{0}), number of entries (n)
y = unlist(x)
mean(y)
length(y)

## T-statistic, using the Wald Test
T_stat = sqrt(length(y)/4.3)*(mean(y) - 4.3)
T_stat

## p-value
pnorm(T_stat)
```

Question 2

The **Augmented Dickey-Fuller (ADF) Test** is a test utilized in time series analysis in order to assess stationarity. There is a function `adf.test()` in the package `tseries` which implements this test. Take a look at `help(adf.test)` for some details.

a. If our objective is to see if there is strong evidence that a time series is stationary, how should the argument `alternative` be set when using `adf.test()`?

This is basically asking about the null / alternative hypotheses. The null hypothesis, H_0, is "the time-series is not stationary"; the alternative hypothesis, H_1, is the "the time series is stationary", and we would like to see if there is strong enough evidence to reject the null and accept the alternative. Hence, the argument `alternative` should be set to `stationary`.

b. Consider the following R commands. This will read in the stock data discussed in lecture, and then run the ADF test on the first stock in the sample. Use this as a starting point to write code to loop over all 1000 stocks and get the p-values for each test. How many of the tests have p-values less than 0.05?

c. Run the p-values found in part (b) through the Benjamini-Hochberg procedure described in lecture. Are any of the series found to be stationary using this approach (again using $\alpha = 0.05$)? Comment on the reason(s) for any differences found.

```
stocksample = read.table("stocksample.txt", header=T,
                         sep="\t", comment.char="")

prop_tot = 0
pvals = c()
for ( i in 1:nrow(stocksample) ) {

  tst = adf.test(as.numeric(log(stocksample[i,5:34])), alternative="s")
  p = tst$p.value
  pvals[i] = p

  if ( p < 0.05 ) {
    prop_tot = prop_tot + 1
  }

}

output_str = paste("The number of p-values less than 0.05 is",
                   prop_tot, ", i.e. the proportion is", prop_tot/nrow(stocksample))
print(output_str)

## Also applying the BH test
print(mean(p.adjust(pvals,method="BH")<0.05))
```

After applying the BH test, we find that the average of adjusted p-vals less than 0.05 is 0.