Homework 6

Ze Yang (zey@andrew.cmu.edu)

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

```r
x = read.table("http://www.stat.cmu.edu/~cschafer/MSCF/PoisHypTest.txt")
x = x$V1
# use the poisson mle
lambda_hat = mean(x)
lambda_0 = 4.3
se = sqrt(lambda_0 / length(x))
# test statistic
wald_t = (lambda_hat - lambda_0)/se

# reject H0 if T < -z_alpha
alpha = c(0.001, 0.05, 0.1)
z = qnorm(1-alpha)
wald_t < -z
```

```
## [1] FALSE FALSE FALSE
```

Therefore, we fail to reject H_0 for $\alpha \leq 0.1$. So the p-value is greater than 0.1. The actual p-value is calculated as

```r
p_val = pnorm(wald_t)
# sanity check: reject H0 if alpha > p, FTR if a < p
test_alpha = c(0, p_val - 10e-6, p_val + 10e-6, 1)
test_z = qnorm(1-test_alpha)
wald_t < -test_z
```

```
## [1] FALSE FALSE  TRUE  TRUE
```

```r
print(p_val)
```

```
## [1] 0.3515102
```

Hence the p-value is 0.3515102.

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()`?

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.

- **Answer (a)**: Since the hypothesis tests are formulated to test **against** the null hypothesis, for our purpose, we want to find stationary stocks. So we should set H_0 : Not stationary, i.e. H_1 : stationary. We will use `alternative='s'`.

```r
stocksample = read.table("stocksample.txt", header=T,
                         sep="\t", comment.char="")
p_values = data.frame(id=rep(0,1000), pval=rep(0,1000))
for (i in 1:1000) {
  adf = adf.test(as.numeric(log(stocksample[i,5:34])), alternative="s")
  p_values$id[i] = i
  p_values$pval[i] = adf$p.value
}

p_values = p_values[order(p_values$pval),]
```

```r
str(p_values)
```

```
## 'data.frame':    1000 obs. of  2 variables:
##  $ id  : num  6 32 58 80 115 160 226 265 281 359 ...
##  $ pval: num  0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 ...
```

```r
nrow(p_values[p_values$pval < 0.05,])
```

```
## [1] 61
```

So there are 61 tests whose p-value is smaller than 0.05, 6.1% of the total.

```r
sum(p.adjust(p_values$pval, method='BH') < 0.05)
```

```
## [1] 0
```

After the adjustment with the Benjamini-Hochberg procedure, there is no p-value that is smaller than 0.05.

Comments

- The different is due to the multiple testing problem. Assume the null hypothesis is true (stock prices are **NOT** stationary), then the p-value follows a uniform distribution. If we sequentially do many tests, we will ultimately see some rejections, i.e. small p-values.

- If we naively apply some constant confidence level, say $\alpha = 0.05$, and regard the tests with $p < 0.05$ as rejects, we are likely to make false positive rejections.
- The Benjamini-Hochberg procedure adjusts the p-values by controling the false positive rates. So it rules out our false discoveries, and we end up seeing zero test having small p-values after this adjustment.
- It's likely that no stock price series is truely stationary.