

DATA-580 Lab 1

Dhun Sheth

2023-09-14

Question 1

```
bpm_data=readRDS(file = "BPM.rds")
```

Part A

```
rng.chisq(bpm_data[1:150,2])[2]
```

```
## $p.value  
## [1] 0.6587231
```

The p-value is greater than 0.05, and not very small, there is little evidence against the uniformity hypothesis and thus the BPM in the first phase follows a uniform distribution.

Part B

```
rng.chisq(bpm_data[,2])[2]
```

```
## $p.value  
## [1] 0.02756615
```

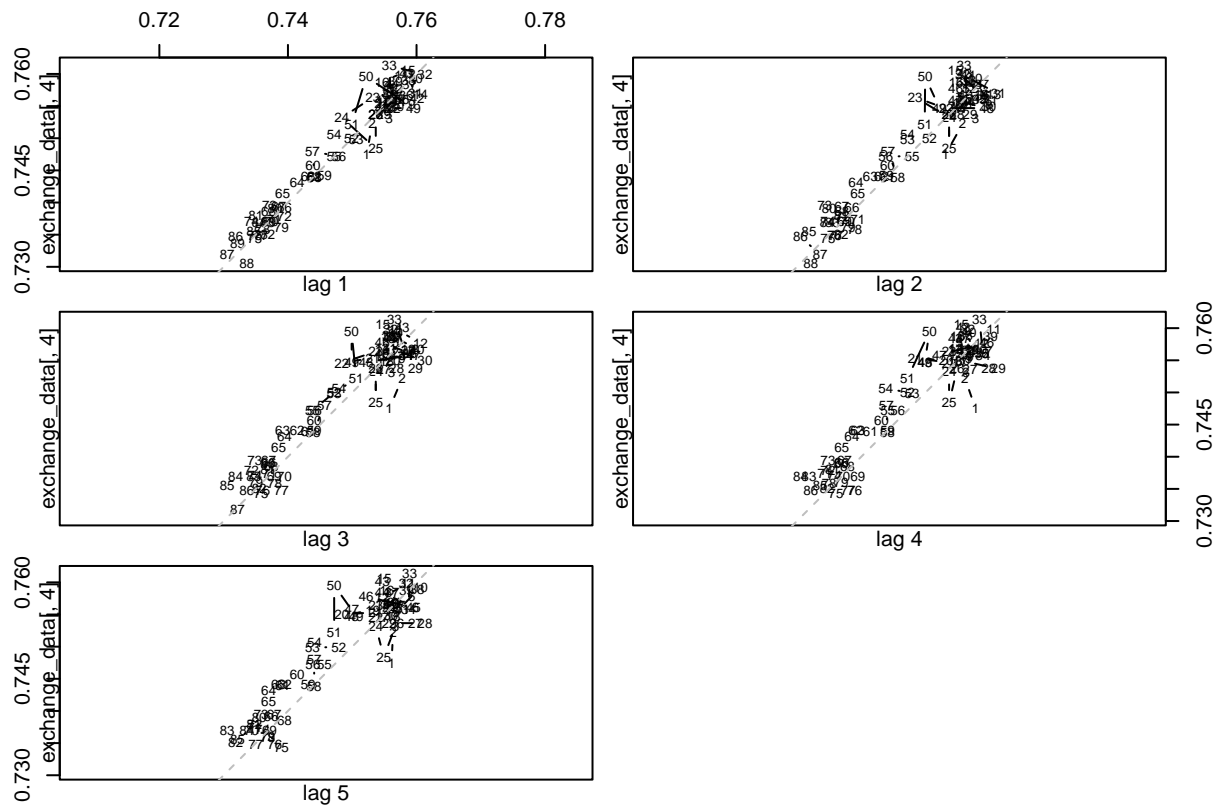
The p-value is less than 0.05 and is very small, there is large evidence against the uniformity hypothesis and thus the entire sequence of BPM does not follow a uniform distribution.

Question 2

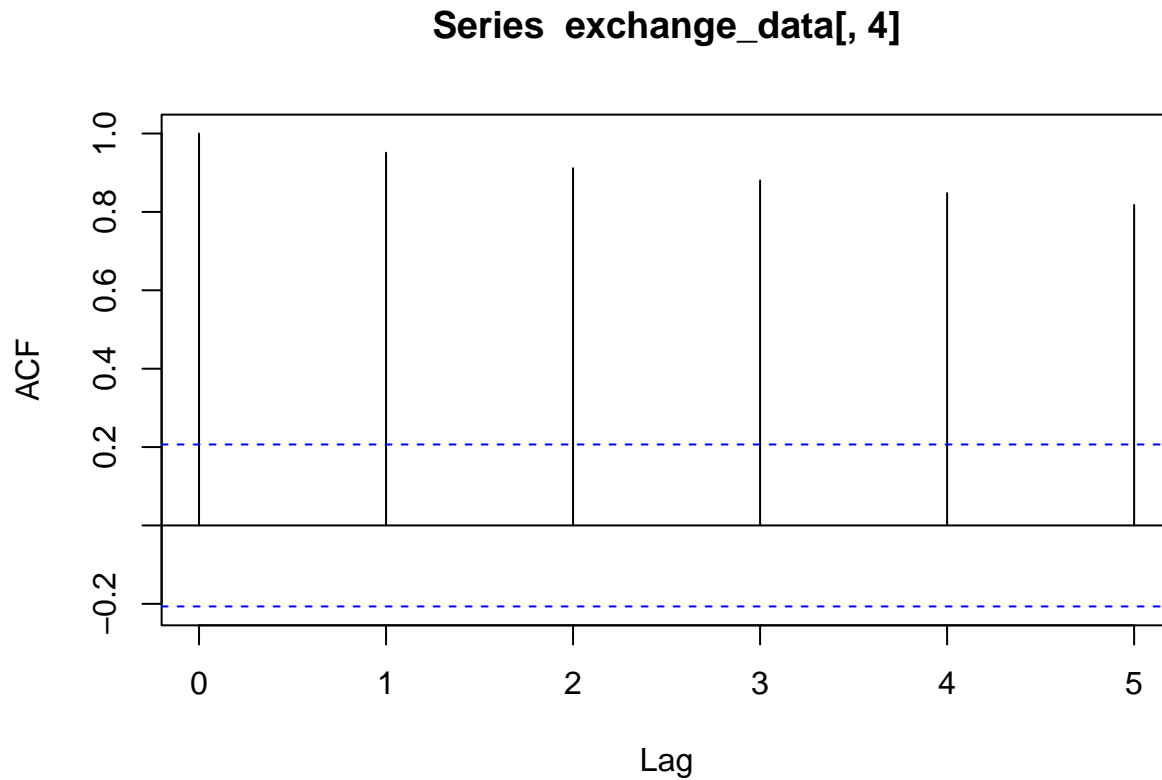
```
exchange_data=readRDS(file = "Exchanges.rds")
```

Part A

```
lag.plot(exchange_data[,4], lags=5)
```



```
acf(exchange_data[,4], 5)
```



Part B

Based on the plot, there are several spikes in the plot which exceed the blue line which indicates dependence. In addition, the lag plots also show a linear dependence.

Question 3

```
dna_data=read.table("DNA.txt", header = FALSE)
```

Part A

```
list_C1_AT = c()
for (k in dna_data[, 2]) {
  n = nchar(k)
  i = 1
  x = rep(0, n)
  while (i < (n - 1)) {
    if (substring(k, i, i) == "A")
      x[i] = 1
    if (substring(k, i + 1, i + 1) == "T")
      x[i + 1] = 1
    i = i + 2
  }
}
```

```

    }
    sigma2 = sum(x^2)/n - (mean(x))^2
    C1 = (sum(x[1:(n - 1)] * x[2:n])/(n - 1) - mean(x[1:(n -
        1)]) * mean(x[2:n]))/sigma2
    list_C1_AT = append(list_C1_AT, C1)
}
list_C1_AT[1:5] # printing seperatly so numbers don't get cut off

```

```
## [1] -0.1254394 -0.1628562 -0.1631519 -0.1254394 -0.1995381
```

```
list_C1_AT[6:length(list_C1_AT)] # printing seperatly so numbers don't get cut off
```

```
## [1] -0.15906818 -0.14754631 0.02576178 -0.14597961 -0.16285624 -0.11157712
```

Part B

```

list_C1_AC = c()
for (k in dna_data[, 2]) {
  n = nchar(k)
  i = 1
  x = rep(0, n)
  while (i < (n - 1)) {
    if (substring(k, i, i) == "A")
      x[i] = 1
    if (substring(k, i + 1, i + 1) == "C")
      x[i + 1] = 1
    i = i + 2
  }
  sigma2 = sum(x^2)/n - (mean(x))^2
  C1 = (sum(x[1:(n - 1)] * x[2:n])/(n - 1) - mean(x[1:(n -
    1)]) * mean(x[2:n]))/sigma2
  list_C1_AC = append(list_C1_AC, C1)
}
list_C1_AC[1:5] # printing seperatly so numbers don't get cut off

```

```
## [1] -0.02532227 -0.01483577 0.18241624 0.01621209 -0.07214815
```

```
list_C1_AC[6:length(list_C1_AC)] # printing seperatly so numbers don't get cut off
```

```
## [1] -0.020691058 -0.029087236 -0.100770302 0.018896365 -0.014835766 -0.002311391
```