Problem Set 3

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

(Due: February 22, in class)

[1] Execute the following lines which create two vectors of random integers which are chosen with replacement from the integers $0, 1, \ldots, 999$. Both vectors have length 250.

```
set.seed(50)
xVec <- sample(0:999, 250, replace=TRUE)
yVec <- sample(0:999, 250, replace=TRUE)
```

Suppose $\mathbf{x} = (x_1, \ldots, x_n)$ denotes the vector `xVec` and $\mathbf{y} = (y_1, \ldots, y_n)$ denotes the vector `yVec`.

a) Create a vector $(y_2 - x_1, \ldots, y_n - x_{n-1})$.
b) Create a vector $(\frac{\sin(y_1)}{\cos(x_2)}, \frac{\sin(y_2)}{\cos(x_3)}, \ldots, \frac{\sin(y_{n-1})}{\cos(x_n)})$.
c) Create a vector $(x_1 + 2x_2 - x_3, x_2 + 2x_3 - x_4 \ldots, x_{n-2} + 2x_{n-1} - x_n)$.
d) Calculate

$$\sum_{i=1}^{n-1} \frac{e^{-x_{i+1}}}{x_i + 10}.$$

[2] This question uses the vectors `xVec` and `yVec` created in the previous question and the functions `sort, order, mean, sqrt, sum` and `abs`.

a) Pick out the values in `yVec` which are > 600.
b) What are the index positions in `yVec` of the values which are > 600?
c) What are the values in `xVec` which correspond to the values in `yVec` which are > 600? (By correspond, we mean at the same index positions.)
d) Create the vector $(|x_1 - \bar{\mathbf{x}}|^{1/2}, \ldots, |x_n - \bar{\mathbf{x}}|^{1/2})$ where $\bar{\mathbf{x}}$ denotes the mean of the vector $\mathbf{x} = (x_1, \ldots, x_n)$.
e) How many values in `yVec` are within 200 of the maximum value of the terms in `yVec`?
f) How many numbers in `xVec` are divisible by 2? (Note that the `modulo` operator is denoted %%.)
g) Sort the numbers in the vector `xVec` in the order of increasing values in `yVec`.
h) Pick out the elements in `yVec` at index positions 1,4,7,10,13,....

[3] Create a 6×10 matrix of random integers chosen from $1, 2, \ldots, 10$ by executing the following two lines of code:

```
set.seed(75)
aMat <- matrix( sample(10, size=60, replace=T), nr=6)
```

a) Find the number of entries in each row which are greater than 4.

b) Which rows contain exactly two occurrences of the number seven?

c) Find those pairs of columns whose total (over both columns) is greater than 75. The answer should be a matrix with two columns; so, for example, the row (1, 2) in the output matrix means that the sum of columns 1 and 2 in the original matrix is greater than 75. Repeating a column is permitted; so, for example, the final output matrix could contain the rows (1, 2), (2, 1) and (2, 2). What if repetitions are not permitted? Then, only (1, 2) from (1, 2), (2, 1) and (2, 2) would be permitted.

[4] a) Write functions `tmpFn1` and `tmpFn2` such that if `xVec` is the vector $(x_1, \ldots, x_n)$, then `tmpFn1(xVec)` returns the vector $(x_1, x_2^2, \ldots, x_n^n)$ and `tmpFn2(xVec)` returns the vector $(x_1, x_2^2/2, \ldots, x_n^n/n)$.

b) Now write a function `tmpFn3` which takes 2 arguments `x` and `n` where `x` is a single number and `n` is a strictly positive integer. The function should return the value of

$$1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots + \frac{x^n}{n}$$

[5] Write a function `tmpFn(xVec)` such that if `xVec` is the vector $\mathbf{x} = (x_1, \ldots, x_n)$ then `tmpFn(xVec)` returns the vector of moving averages:

$$\frac{x_1 + x_2 + x_3}{3}, \frac{x_2 + x_3 + x_4}{3}, \ldots, \frac{x_{n-2} + x_{n-1} + x_n}{3}$$

Try out your function; for example, try `tmpFn( c(1:5,6:1) )`.

[6] Consider the continuous function

$$f(x) = \begin{cases} x^2 + 2x + 3 & x < 0 \\ x + 3 & 0 \le x < 2 \\ x^2 + 4x - 7 & 2 \le x \end{cases}$$

Write a function `tmpFn` which takes a single argument `xVec`. The function should return the vector of values of the function $f(x)$ evaluated at the values in `xVec`.

[7] Given a vector $\mathbf{x} = (x_1, \ldots, x_n)$, the *sample autocorrelation of lag k* is defined to be

$$r_k = \frac{\sum_{i=k+1}^{n} (x_i - \bar{\mathbf{x}})(x_{i-k} - \bar{\mathbf{x}})}{\sum_{i=1}^{n} (x_i - \bar{\mathbf{x}})^2}$$

a) Write a function `tmpFn1(xVec)` which takes a single argument `xVec` which is a vector and returns a scalar r_1. In particular, for the vector $\mathbf{X} = (2, 5, 8, \ldots, 53, 56)$ compute `tmpFn1(X)`.

b) Write a function `tmpFn2(xVec)` which takes a single argument `xVec` which is a vector and returns a `list` of two values: r_1 and r_2. In particular, for the vector $\mathbf{X} = (2, 5, 8, \ldots, 53, 56)$ compute `tmpFn2(X)`.

[8] Modify the function `primeFinder` in the lecture notes by using `while` loop instead of `for` in the body of the function.

[9] For two consecutive primes x_n and x_{n+1}, if $x_{n+1} = x_n + 2$, they are called *twin primes*. For example, (3,5), (5,7), (11,13) are twin primes. Write a function `twinPrimeFinder` to find all twin primes less than a given number n. (Look at the function `primeFinder` in the lecture notes.)

[10] This exercise contains modifications of Newton's method mentioned in the lectures notes.

a) Modify the loop `fNewton` in the lecture notes to find one of the zeros of

$$\cos(x) = e^x$$

b) Put this loop into function `fNewtonFun(x0)` which will return a root starting from an initial point `x0` if the method converges to a root and `NA` otherwise.