STAT 406: HW1

- All computer code should be written using the language R. Type ALL your code into one PLAIN Text format file. Plain text format is available by default in R. Please do not use Microsoft Word .doc format or .rtf format of .pdf format. Inside your plain text file, make sure you identify each problem in a comment placed at the beginning of the problem. The file name should match your name as in 'JohnDoe.R'. Submit your R code file online (under Assignments) at or before the due date, and hand in a hard copy of the code with any additional material to your lab.
- I recommend that before submitting your homework, you also create a new directory and run your R code, to make sure that it is self-contained and runs as you intended.

1. Suppose that U_1 and U_2 are independent random variables with uniform distribution $\mathsf{U}(0,1)$. From U_1, U_2 we build the random variable Z as follows
$$Z = \sqrt{-2\log(U_1)}\cos(2\pi U_2)\,.$$
   (a) Write a code that build a vector (called vecZ) of length $n = 1000$, each component of which is a realization of a random variable with the same distribution as Z. **Note**: for/while loops not allowed. Use the function runif to generate uniform random numbers.
   (b) Write appropriate R statements to find the proportion of the elements of vecZ that fall: (a) below 0; (b) between 0 and 1; (c) above 1.

2. For integer $n \geq 1$, and $x \in [0,1]$, define
$$B_n(x) = \sum_{k=0}^{n} \left|\sin\left(\frac{2\pi k}{n}\right)\right| \binom{n}{k} x^k (1-x)^{n-k}. \tag{1}$$

In the above formula, $\binom{n}{k}$ denotes the combinatorial number "n choose k" which can be computed in R using the function choose. Without using any for/while loop, compute $B_n(x)$ for $x = 0.2$, and $n = 20$, taking full advantage of the vectorized nature of R.

3. Daily precipitation has been collected for 56 weather stations in the Colorado Front Range and made available as a R list contained in the binary R file *FrontRange.RData*. Use the command

```
load(FrontRange.RData)
```

to load the data set (called **FR**) in R. The data set **FR** is a list with 4 components "precip", "time", "info" and "Stot". The components "precip" and "times" are also lists, whereas "info" is a data frame and "Stot" is a vector. The list "precip" has 56 components, one for each station. *FR$precip[[i]]* thus has all the daily rainfalls recorded for the i-th station. In a similar structure, the component "time" of **FR** hold the dates at which the rainfalls are recorded. The component "info" is a data frame with various information on the stations (longitude, latitude, elevation, number of records, plus some other variables that are not so important for us here). Finally, the component "Stot" is a vector that contains the mean summer precipitation for each station.

   (a) Find the station at the highest elevation and the station at the lowest elevation. Plot the daily rainfalls over time recorded by these two stations.

   (b) For each of the two stations identified above, what is the proportion of days with 10mm of rain or more?

   (c) Write code to find for a given station (say station 10), the date of the record highest amount of precipitation.