Loops in R

There are primarily two loop structures used in R; the **for** and the **while** loop structure. The structure is very similar to other languages.

- **for** loop has the following structure:

  for (*index_name* **in** *vector_of_iterates*)

  {

  ....

  ....

  }

  As an example, consider the following for loop structure

  natural_numbers = seq(1,10,2)

  sum = 0

  for(i in natural_numbers)

  {

  sum = sum + i;

  }

  What would sum be after the loop terminates?

- The **while** loop has the following structure:

  while (*complement of the exit condition*)

  {

  ....

  ......

  }

  In other words, a while loop runs as long as a certain set of defined conditions are met. Once they are violated, the loop terminates. (Recommended specifically for cases when the number of iterations is not known *apriori*.

Functions in R

function has the following structure:
required_function=**function**(input1,input2,...,input$n$) { ...
...
...
return(output)
}
As an example, consider the following:
$sol = function(a)$
{
$b = 0$
$for(i\ in\ 1 : length(a))\{$
$b = b + a[i]$
}
$return(b)$
}

$a = seq(1, 50, 2)$
$sol(a)$

What will the above function return?

Exercise 1: Let $X_1, X_2, ..., X_n$ be i.i.d samples drawn from a normal distribution with mean 0 and variance 1. Let $Y_1, Y_2, ..., Y_n$ be i.i.d samples drawn from a Cauchy distribution with location parameter 0 and scale parameter 1. Define $\bar{X}_n$ and $\bar{Y}_n$ to be sample mean of X_i's and Y_i's respectively. Define $\tilde{X}_n$ and $\tilde{Y}_n$ to be sample median of X_i's and Y_i's respectively.

1. **[R]** Plot the above four statistics and observe their asymptotic behaviour as $n \to \infty$.
2. **[R]** Empirically comment about the convergence of each of the above four statistics.

(To show multiple plots in the same graphic window, search and study for the R function **split.screen()** and **screen()** functions. You can also look for **mfrow** in **par**.)

Exercise 2: A virus

Bindu is a student in a class of n students. Each student in the class has stored in his/her computer the email addresses of all other students (but not his/her own). Assume that the students have not stored any other email-addresses of people outside of the class. Bindu's computer is infected with a virus. The virus selects exactly one address randomly from the addresses stored on the computer and spreads to the computer of that student. Then it tries to spread in the same fashion from the newly infected computer. If at any stage, the next computer that the virus selects is already infected, the virus stops spreading any further. We are interested in finding how many students are expected to have their computers infected (including Bindu) after the virus has stopped spreading.

(a) **[R]** Explain clearly how you can simulate this situation. Your answer should mention what type of loops and/or arrays will be used and why your logic is correct.

(b) Implement a `simvirus(n)` which takes n (the number of students) as an input and returns the number of computers infected in a single simulated run.

(c) **[R]** Run your simulation keeping $n = 20$ for a sufficiently large number of times and report the mean and the variance (and the number of runs used).

(d) Plot a histogram of the frequency of different values of k.

(e) **[R]** Comment on the histogram and why you think the distribution shown makes sense.