Biostatistics 615/815
Problem Set 2
Due Wednesday, September 29

Floating Point Precision

1. This problem aims to explore the accuracy of floating point representations in your system. You should investigate the float and double data types in C as well as the default floating point representation in R. Write a program that calculates the following three quantities:

    a. The smallest positive number ε such that $1.0 + \varepsilon > 1.0$.
    b. The smallest positive number ε* such that $1.0 - \varepsilon^* < 1.0$.
    c. The smallest positive number that can be stored in a variable and is > 0.0.

2. Care must be taken when evaluating certain mathematical expressions on a computer, as rounding errors can build up to become quite large. One well known example relates to the golden ratio, φ. This number is:

$$\phi = \frac{\sqrt{5}-1}{2} = 0.61803398$$

The number has the useful property that:

$$\phi^n = \phi^{n-1}\phi \qquad (1)$$

$$\phi^n = \phi^{n-2} - \phi^{n-1} \qquad (2)$$

Thus, successive powers of φ can be conveniently calculated using either expression (1) or (2).

Starting with $\varphi^0 = 1$ and φ^1 defined above, write an R program that fills an array with successive powers of φ using expression (1) and another array using expression (2). Plot both sets of results for $\varphi^{n = 1..80}$. Which one appears more accurate? Why?