Computational Statistics Lab 1 Report Group 6

Jaskirat S Marar (jasma356), Filip Berndtsson (filbe354), Dinuke Jayaweera (dinja628), Raja Uzair Saee

11/9/2021

Problem 1

1. Check the results of the snippet. Comment on whats going on

First Snippet:

```
x1 <-  1/3
x2 <- 1/4
if (x1 - x2 == 1/12) {
  print("Subtraction is correct")
  } else {
    print("Subtraction is wrong")
  }
```

```
## [1] "Subtraction is wrong"
```

If we examine the problem by each variable separately we see that the computed values are as follows:

```
x1 <-  1/3
x2 <- 1/4
x3 <- 1/12

x1
```

```
## [1] 0.33333333333333331
```

```
x2
```

```
## [1] 0.25
```

```
x3
```

```
## [1] 0.083333333333333329
```

```
x1-x2
```

```
## [1] 0.083333333333333315
```

We can see that since x1 has recurring decimals, it gets stored with a rounding error. Whereas, x2 is a number with a denominator that is a power of 2 and hence it doesn't have a mantissa and hence no rounding error. In the case for x3 i.e 1/12, it is the same as x1 where it has a recurring decimal case and get stored with a rounding error. When we perform the subtraction operation of x1-x2, the dedicated higher precision register where this operation takes place first lines up the decimals and performs the operations and then only after the operation is complete, it performs the rounding which leads to different values getting carried forward due to rounding. Hence we see the difference in the results of x1-x2 and 1/12.

Second Snippet:

```r
x1 <- 1
x2 <- 1/2
if (x1 - x2 == 1/2) {
  print("Subtraction is correct")
} else {
    print("Subtraction is wrong")
  }
```

```
## [1] "Subtraction is correct"
```

We don't run into the same problem here as the first snippet because both x1 and x2 can be expressed as powers of 2 and hence dont encounter any rounding error when being stored in the machine. Therefore, here the result is accurate.

2. If there are any problems suggest solutions

One possible way to correct the error in the first snippet can be to sue the equivalent algebraic expression (x - y)(x + y) = (x^2 - y^2) Sometimes this kind of an algebraic trick can help to reduce the loss of significance. We can see the results of this change below

```r
x1 <-  1/3
x2 <- 1/4
x_res <- (x1^2 - x2^2)/(x1+x2)
if (x_res == 1/12) {
  print("Subtraction is correct")
  } else {
    print("Subtraction is wrong")
  }
```

```
## [1] "Subtraction is correct"
```

Problem 2

1. Write your own R function to calculate the derivative of f(x) = x in this way with ϵ = 10^{-15}

2. Evaluate your derivative function at x = 1 and x = 100000

3. What values did you obtain? What are the true values? Explain the reasons behind the discovered differences

Problem 3

1. Write your own R function, myvar, to estimate the variance in this way

2. Generate a vector x = (x$_1$,. . . , x$_{10000}$) with 10000 random numbers with mean 10^8 and variance 1

3. For each subset X$_i$ = {x$_1$, . . . , x$_i$}, i = 1,. . . , 10000 compute the difference Y$_i$ = myvar(X$_i$)-var(X$_i$), where var(X$_i$) is the standard variance estimation function in R. Plot the dependence Y$_i$ on *i*. Draw conclusions from this plot. How well does your function work? Can you explain the behaviour?

4. How can you better implement a variance estimator? Find and implement a formula that will give the same results as var()?

Problem 4

Consider the three below R expressions for computing the binomial coefficient. They all use the prod() function, which computes the product of all the elements of the vector passed to it.

A) prod (1 : n) / (prod (1 : k) * prod (1 : (n-k)))

B) prod((k+1) : n) / prod (1 : (n-k))

C) prod (((k+1) : n) / (1 : (n-k)))

1. Even if overflow and underflow would not occur these expressions will not work correctly for all values of n and k. Explain what is the problem in A, B and C respectively

2. In mathematical formulae one should suspect overflow to occur when parameters, here n and k, are large. Experiment numerically with the code of A, B and C, for different values of n and k to see whether overflow occurs. Graphically present the results of your experiments.

3. Which of the three expressions have the overflow problem? Explain why.