

Computer Lab 1

Computational Statistics

Linköpings Universitet, IDA, Statistik

2022 XI 09

Kurskod och namn:	732A90 Computational Statistics
Datum:	2022 XI 08—2022 XI 15 (lab session 9 XI 2022)
Delmomentsansvarig:	Krzysztof Bartoszek, Bayu Brahmantio, Jaskirat Marar, Shashi Nagarajan
Instruktioner:	<p>This computer laboratory is part of the examination for the Computational Statistics course</p> <p>Create a group report, (that is directly presentable, if you are a presenting group), on the solutions to the lab as a .PDF file.</p> <p>Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.</p> <p>All R code should be included as an appendix into your report.</p> <p>A typical lab report should 2-4 pages of text plus some amount of figures plus appendix with codes.</p> <p>In the report reference ALL consulted sources and disclose ALL collaborations.</p> <p>The report should be handed in via LISAM</p> <p>(or alternatively in case of problems e-mailed to bayu_dot_brahmantio_at sign_liu_dot_se, jasma356_at sign_student_dot_liu_dot_se, shashi_dot_nagarajan_at sign_@liu_dot_se, or krzysztof_dot_bartoszek_at sign_liu_dot_se),</p> <p>by 23:59 15 November 2022 at latest.</p> <p>Notice there is a deadline for corrections 23:59 22 January 2023 and a final deadline of 23:59 12 February 2023 after which no submissions nor corrections will be considered and you will have to redo the missing labs next year.</p> <p>The seminar for this lab will take place 29 November 2022.</p> <p>The report has to be written in English.</p>

Question 1: Be careful when comparing

Consider the following two R code snippets

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

and

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

1. Check the results of the snippets. Comment what is going on.
2. If there are any problems, suggest improvements.

Question 2: Derivative

From the definition of a derivative a popular way of computing it at a point x is to use a small ϵ and the formula

$$f'(x) = \frac{f(x + \epsilon) - f(x)}{\epsilon}.$$

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

Question 3: Variance

A known formula for estimating the variance based on a vector of n observations is

$$\text{Var}(\vec{x}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right)$$

1. Write your own R function, `myvar`, to estimate the variance in this way.
2. Generate a vector $x = (x_1, \dots, x_{10000})$ with 10000 random numbers with mean 10^8 and variance 1.
3. For each subset $X_i = \{x_1, \dots, x_i\}$, $i = 1, \dots, 10000$ compute the difference $Y_i = \text{myvar}(X_i) - \text{var}(X_i)$, 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()`?

Question 4: Binomial coefficient

The binomial coefficient “ n choose k ” is defined as

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} = \frac{(k+1)(k+2) \cdots (n-1)n}{(n-k)!},$$

where n and k are an arbitrary pair of integers satisfying $0 \leq k \leq n$. 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.