

# Computer Lab 6

## Computational Statistics

Linköpings Universitet, IDA, Statistik

2022 XII 14

---

Kurskod och namn:	732A90 Computational Statistics
Datum:	2022 XII 12—2022 XII 20 (lab session 14 XII 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 <b>.PDF</b> 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><b>All R code should be included as an appendix into your report.</b></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 <b>ALL</b> consulted sources and disclose <b>ALL</b> 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 <b>23:59 20 December 2022</b> at latest.</p> <p>Notice there is a deadline for corrections <b>23:59 22 January 2023</b> and a final deadline of <b>23:59 12 February 2023</b> 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 <b>13 December 2022</b>.</p> <p>The report has to be written in English.</p>

---

## Question 1: Genetic algorithm

In this assignment, you will try to perform one-dimensional maximization with the help of a genetic algorithm.

1. Define the function

$$f(x) := \frac{x^2}{e^x} - 2 \exp(-(9 \sin x)/(x^2 + x + 1))$$

2. Define the function `crossover()`: for two scalars  $x$  and  $y$  it returns their “kid” as  $(x+y)/2$ .
3. Define the function `mutate()` that for a scalar  $x$  returns the result of the integer division  $x^2 \bmod 30$ . (Operation `mod` is denoted in R as `%%`).
4. Write a function that depends on the parameters `maxiter` and `mutprob` and:
  - (a) Plots function  $f$  in the range from 0 to 30. Do you see any maximum value?
  - (b) Defines an initial population for the genetic algorithm as  $X = (0, 5, 10, 15, \dots, 30)$ .
  - (c) Computes vector `Values` that contains the function values for each population point.
  - (d) Performs `maxiter` iterations where at each iteration
    - i. Two indexes are randomly sampled from the current population, they are further used as parents (use `sample()`).
    - ii. One index with the smallest objective function is selected from the current population, the point is referred to as victim (use `order()`).
    - iii. Parents are used to produce a new kid by crossover. Mutate this kid with probability `mutprob` (use `crossover()`, `mutate()`).
    - iv. The victim is replaced by the kid in the population and the vector `Values` is updated.
    - v. The current maximal value of the objective function is saved.
  - (e) Add the final observations to the current plot in another colour.
5. Run your code with different combinations of `maxiter`= 10, 100 and `mutprob`= 0.1, 0.5, 0.9. Observe the initial population and final population. Conclusions?

## Question 2: EM algorithm

The data file `physical.csv` describes a behavior of two related physical processes  $Y = Y(X)$  and  $Z = Z(X)$ .

1. Make a time series plot describing dependence of  $Z$  and  $Y$  versus  $X$ . Does it seem that two processes are related to each other? What can you say about the variation of the response values with respect to  $X$ ?
2. Note that there are some missing values of  $Z$  in the data which implies problems in estimating models by maximum likelihood. Use the following model

$$Y_i \sim \exp(X_i/\lambda), \quad Z_i \sim \exp(X_i/(2\lambda))$$

where  $\lambda$  is some unknown parameter.

**The goal is to derive an EM algorithm that estimates  $\lambda$ .**

3. Implement this algorithm in R, use  $\lambda_0 = 100$  and convergence criterion “stop if the change in  $\lambda$  is less than 0.001”. What is the optimal  $\lambda$  and how many iterations were required to compute it?
4. Plot  $E[Y]$  and  $E[Z]$  versus  $X$  in the same plot as  $Y$  and  $Z$  versus  $X$ . Comment whether the computed  $\lambda$  seems to be reasonable.