Computational Statistics
Exam, Fall 2023

Department of Computer and Information Science (IDA), Linköping University

May 13, 2024, 8:00-13:00

Course:	732A90 Computational Statistics
Teachers:	Krzysztof Bartoszek, Frank Miller (examiner)
Allowed aids:	Printed course books, 100 page computer document
Provided aids:	material in zip file exam_material_732A90.zip
Grades:	A = [18, 20] points, B = [16, 18) points, C = [14, 16) points, D = [12, 14) points, E = [10, 12) points, F = [0, 10) points.
Instructions:	Provide a detailed report that includes plots, conclusions and interpretations. If you are unable to include a plot in your solution file clearly indicate the section of R code that generates it. Give motivated answers to the questions. If an answer is not motivated, the points are reduced. Provide all necessary codes in an appendix. In a number of questions you are asked to do plots. Make sure that they are informative, have correctly labelled axes, informative axes limits and are correctly described. Points may be deducted for poorly done graphs. If you have problems with creating a pdf you may submit your solutions in text files with unambiguous references to graphics and code that are saved in separate files. There are TWO assignments (with sub–questions) to solve. Provide a separate solution file for each assignment. Include all R code that was used to obtain your answers in your solution files. Make sure it is clear which code section corresponds to which question. If you also need to provide some hand–written derivations please number each page according to the pattern: Question number . page in question number, i.e. Q1.1, Q1.2, . . ., Q2.1, Q2.2, ... Name your solution files as: [your exam account id] [own file description].[format]

Note: If you are not able to solve a part of a question, you can anyway try to show how you would solve the subsequent parts with explaining and providing code examples and you might receive partial points.

Assignment 1 (10p)

Consider the function

$$g(x, y) = -x^2 - x^2y^2 - 2xy + 2x + 2.$$

The gradient and the Hessian matrix are

$$g'(x, y) = \begin{pmatrix} -2x - 2xy^2 - 2y + 2 \\ -2x^2y - 2x \end{pmatrix}, \qquad g''(x, y) = \begin{pmatrix} -2 - 2y^2 & -4xy - 2 \\ -4xy - 2 & -2x^2 \end{pmatrix},$$

respectively. It is desired to determine the point (x, y), $x, y \in [-3, 3]$, where the function is maximized.

Question 1.1 (4p)

Write an own algorithm based on the Newton method in order to find a point where the gradient is 0. Extend your code such that an output is printed saying if the point found is a local maximum, local minimum, or a saddle point of g.

Question 1.2 (3p)

Use the three points $(x, y) = (2, 0), (-1, -2), (0, 1)$ as starting values. Describe what happens when you run your algorithm for each of those starting values.

Question 1.3 (1p)

Describe briefly (about 2-3 lines) an idea how to extend the algorithm such that it searches a local maximum by continuing if the point found is a local minimum or a saddle point. You need not to program this.

Question 1.4 (2p)

What would be the advantages and disadvantages when you would run a steepest ascent algorithm instead of the Newton algorithm?

Assignment 2 (10p)

Consider the following density:

$$f(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{5}{6} \cdot \exp(-x) \cdot (1 + \cos(2x)), & \text{if } x \geq 0. \end{cases}$$

We are interested to generate draws of a random variable X with this density using rejection sampling.

Question 2.1 (3p)

Choose two envelopes: a good envelope $e(x)$ and a bad envelope $e_{\text{bad}}(x)$ for rejection sampling. Motivate why you think that they are good or bad, respectively.

Question 2.2 (1p)

Write a function in `R` for the density $f(x)$ and the two envelopes $e(x), e_{\text{bad}}(x)$ and plot the functions over the range $0 \leq x \leq 5$.

Question 2.3 (3p)

Program a random generator for X using rejection sampling for the chosen envelope $e(x)$. Generate 10000 random variables and plot a histogram.

Question 2.4 (3p)

Based on the sample generated in Part 2.3, determine estimates for $P(X > 1)$ and $E(X)$. Determine some measure of uncertainty for these estimates.

If you failed to generate the sample in Part 2.3, you can do this Part 2.4 by generating such a sample for the exponential distribution using build-in functions in `R` *and calculating the things here based on this (wrong) sample (but not based on build-in functions for the exponential distribution), which still might give you the points for this part.*